GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:55:00 ; Search time 11.2849 Seconds
(without alignments)
5501.769 Million cell updates/sec

Title: US-10-676-079-6
Perfect score: 23
Sequence: 1 ttcgatcccaagaaggaatcaac 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23	100.0	23	20	AAX35646	PCR primer used to
	2	23	100.0	23	21	AAA75049	PCR primer HPU335
	3	23	100.0	23	21	AAA75063	PCR primer Hpu-335
	4	23	100.0	23	21	AAZ33293	Human heparanase P
	5	23	100.0	1584	24	ABL40753	Chicken signal pep
	6	23	100.0	1593	20	AAZ11236	Human pre-prohepar
	7	23	100.0	1669	25	ABZ22816	Human heparanase e
	8	23	100.0	1713	20	AAX37259	Human heparanase e
	9	23	100.0	1721	20	AAX35648	cDNA encoding a hu
	10	23	100.0	1721	21	AAA75051	cDNA encoding a hu
	11	23	100.0	1721	21	AAZ39195	Human heparanase e
	12	23	100.0	1721	21	AAZ33290	Human heparanase n
	13	23	100.0	1721	22	AAA91112	Human heparanase,
	14	23	100.0	1722	22	AAF93788	Human cDNA encodin
	15	23	100.0	1723	20	AAX37260	Seq ID No: 14 of W
	16	23	100.0	1724	22	AAH20940	Human heparanase i
	17	23	100.0	1899	20	AAX35650	cDNA encoding a hu
	18	23	100.0	1899	21	AAA75053	cDNA encoding a hu
	19	23	100.0	3726	20	AAX86671	cDNA encoding a hu
	20	23	100.0	44848	21	AAA75080	Nucleotide sequenc
	21	22	95.7	474	22	AAF93984	Primer specific fo
c	22	17.2	74.8	2884	16	AAQ84613	Rat AT2 receptor c
	23	17.2	74.8	159400	24	ABQ88126	Human osteoblast d
c	24	17.2	74.8	160271	22	AAF85750	Bipolar affective
c	25	17.2	74.8	160271	22	AAF85756	Human chromosome 1
c	26	17.2	74.8	160271	22	AAS04858	Human chromosome 1
c	27	17.2	74.8	160271	22	AAS04864	Human chromosome 1
c	28	17.2	74.8	160271	22	AAS06667	Human chromosome 1
c	29	17.2	74.8	160271	22	AAH40997	160kb fragment of
c	30	17.2	74.8	160271	22	AAH23764	Human chromosome 1
c	31	17.2	74.8	160271	22	AAF85116	Nucleotide sequenc
c	32	16.8	73.0	343	21	AAC17423	Human secreted pro
c	33	16.8	73.0	367	24	ABN15848	Human ORFX polynuc
c	34	16.8	73.0	577	22	ABA63814	Human foetal liver
c	35	16.8	73.0	577	22	ABA30999	Probe #9465 for ge
c	36	16.8	73.0	577	22	AAK12323	Human brain expres
c	37	16.8	73.0	577	22	AAK38044	Human bone marrow
c	38	16.8	73.0	577	22	AAI18816	Probe #8749 for ge
c	39	16.8	73.0	577	22	AAI43938	Probe #12624 used
c	40	16.8	73.0	577	23	ABS37665	Human liver single
c	41	16.8	73.0	577	24	ABS12053	Human genome-deriv
c	42	16.8	73.0	588	22	AAH08812	Human cDNA clone (
	43	16.8	73.0	1731	22	AAH13668	Human cDNA sequenc
	44	16.8	73.0	2747	23	AAS64307	DNA encoding novel
	45	16.8	73.0	2747	23	AAS64928	DNA encoding novel

ALIGNMENTS

RESULT 1
AAX35646
ID AAX35646 standard; DNA; 23 BP.
XX
AC AAX35646;
XX
DT 09-JUL-1999 (first entry)
XX
DE PCR primer used to amplify human hp3 cDNA.
XX
KW Heparanase; hp; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW PCR primer; ss.
XX
OS Synthetic.
XX
PN WO9911798-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US17954.
XX
PR 02-JUL-1998; 98US-0109386.
PR 02-SEP-1997; 97US-0922170.
XX
PA (FRIE/) FRIEDMAN M M.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Feinstein E, Pecker I, Vlodavsky I;
XX
DR WPI; 1999-302255/25.
XX
PT New human polynucleotide useful for treating angiogenesis,
PT restenosis, and inflammation
XX
PS Example 1; Page 23; 63pp; English.
XX
CC The specification describes a polypeptide having heparanase (hp)
CC activity. The recombinant protein is used as a modulator of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections
CC or disintegration of neurodegenerative plaques. Heparanase may be
CC useful for conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC and renal failure in biopsy specimens, plasma samples, and body fluids.
CC PCR primers AAX35646-47 were used to amplify hp3 cDNA, in the course of
CC the invention.
XX
SQ Sequence 23 BP; 9 A; 6 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 23; DB 20; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCGATCCCAAGAAGGAATCAAC 23
||||||||||||||||||||||
Db 1 TTCGATCCCAAGAAGGAATCAAC 23

RESULT 2
AAA75049
ID AAA75049 standard; DNA; 23 BP.
XX
AC AAA75049;
XX
DT 15-JAN-2001 (first entry)
XX
DE PCR primer HPU335 used to amplify human cDNA encoding heparanase.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US03542.
XX
PR 01-MAR-1999; 99US-0258892.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
DR WPI; 2000-579289/54.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumour, inflammation, autoimmunity, neurodegenerative diseases -
XX
PS Disclosure; Page 44; 152pp; English.
XX
CC The present PCR primer was used to amplify a human cDNA sequence,
CC which encoded a protein with heparanase catalytic activity. The
CC heparanase (hpa) polynucleotide is useful in gene therapy, particularly
CC in treating tumour, inflammation or autoimmunity. Particularly, the
CC polynucleotide is useful in modulating the bioavailability of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors (e.g. bFGF) and cytokines (e.g. interleukin (IL)-8),
CC cell interaction with plasma lipoproteins, cellular susceptibility to
CC certain viral and some bacterial and protozoa infections, or
CC disintegration of neurodegenerative plaques. The polynucleotide is
CC also useful in wound healing (e.g. thermal, chemical or radiation burns),
CC and in the treatment of angiogenesis, restenosis, atherosclerosis,
CC inflammation, neurodegenerative diseases (Gerstmann-Straussler Syndrome
CC or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa
CC infections.
XX
SQ Sequence 23 BP; 9 A; 6 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCGATCCCAAGAAGGAATCAAC 23
||||||||||||||||||||||
Db 1 TTCGATCCCAAGAAGGAATCAAC 23

RESULT 3
AAA75063
ID AAA75063 standard; DNA; 23 BP.
XX
AC AAA75063;
XX
DT 15-JAN-2001 (first entry)
XX
DE PCR primer Hpu-335 used to amplify human cDNA encoding heparanase.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US03542.
XX
PR 01-MAR-1999; 99US-0258892.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
DR WPI; 2000-579289/54.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumour, inflammation, autoimmunity, neurodegenerative diseases -
XX
PS Disclosure; Page 45; 152pp; English.
XX
CC The present PCR primer was used to amplify a human cDNA sequence,
CC which encoded a protein with heparanase catalytic activity. The

CC heparanase (hpa) polynucleotide is useful in gene therapy, particularly
CC in treating tumour, inflammation or autoimmunity. Particularly, the
CC polynucleotide is useful in modulating the bioavailability of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors (e.g. bFGF) and cytokines (e.g. interleukin (IL)-8),
CC cell interaction with plasma lipoproteins, cellular susceptibility to
CC certain viral and some bacterial and protozoa infections, or
CC disintegration of neurodegenerative plaques. The polynucleotide is
CC also useful in wound healing (e.g. thermal, chemical or radiation burns),
CC and in the treatment of angiogenesis, restenosis, atherosclerosis,
CC inflammation, neurodegenerative diseases (Gerstmann-Straussler Syndrome
CC or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa
CC infections.
XX
SQ Sequence 23 BP; 9 A; 6 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCGATCCCAAGAAGGAATCAAC 23
||||||||||||||||||||||
Db 1 TTCGATCCCAAGAAGGAATCAAC 23

RESULT 4
AAZ33293
ID AAZ33293 standard; DNA; 23 BP.
XX
AC AAZ33293;
XX
DT 21-FEB-2000 (first entry)
XX
DE Human heparanase PCR primer Hpu-355 SEQ ID NO:6.
XX
KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW autoimmune disease; anticancer; kidney disease; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09255.
XX
PR 01-MAY-1998; 98US-0071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
XX
PT Heparanase-specific molecular probes useful for diagnosis and
PT treatment, e.g. of tumors, and for targeted drug delivery -
XX
PS Example; Page 30; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and
CC anti-heparanase antibodies are used to detect or quantify the expression
CC of heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary,
CC cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC syndrome, sepsis and inflammatory or autoimmune disease), for targeted
CC drug delivery (e.g. of anticancer agents) and as research reagents.
CC The present sequence represents a PCR primer for human heparanase, which
CC is used in an example from the present invention.
XX
SQ Sequence 23 BP; 9 A; 6 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCGATCCCAAGAAGGAATCAAC 23
||||||||||||||||||||||
Db 1 TTCGATCCCAAGAAGGAATCAAC 23

RESULT 5
ABL40753
ID ABL40753 standard; cDNA; 1584 BP.
XX
AC ABL40753;
XX
DT 03-JUL-2002 (first entry)
XX
DE Chicken signal peptide/human heparanase chimeric cDNA.
XX
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric; ss.
XX
OS Synthetic.
OS Gallus gallus.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1584
FT /*tag= a
FT /product= "chimeric chicken-human heparanase"
FT sig_peptide 1..57
FT /*tag= b
FT /note= "chicken heparanase signal peptide"
FT mat_peptide 58..1581
FT /*tag= c
FT /note= "human mature heparanase"
XX
PN US2002034810-A1.
XX
PD 21-MAR-2002.
XX
PF 16-AUG-2001; 2001US-0930218.
XX
PR 20-SEP-2000; 2000US-0666390.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
XX
DR WPI; 2002-338926/37.
DR P-PSDB; ABB07815.
XX
PT Nucleic acid encoding avian and reptile heparanase polypeptide is
PT useful to treat various heparin-related disorders and the signal
PT peptide is useful in production of membrane-targeted or secreted
PT recombinant proteins -
XX
PS Disclosure; Page 24-25; 39pp; English.
XX
CC The invention relates to an isolated avian and reptile nucleic acid,
CC encoding a polypeptide with heparanase catalytic activity. The signal
CC peptide of the nucleic acid can be used to express membrane-associated or
CC secreted proteins in heterologous expression systems. The encoded

```
CC   polypeptides can be used to prevent tumour angiogenesis, metastasis and
CC   invasion, and to intervene with pathologies associated with impaired
CC   heparin-binding growth factors, cellular responses to heparin-binding
CC   growth factors and cytokines, cell interaction with plasma lipoproteins,
CC   cellular susceptibility to viral, protozoa and bacterial infections or
CC   disintegration of neurodegenerative plaques. The present sequence
CC   represents a chicken signal peptide/human heparanase chimeric cDNA
CC   sequence.
XX
SQ   Sequence 1584 BP; 424 A; 361 C; 373 G; 426 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 24;  Length 1584;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        262 TTCGATCCCAAGAAGGAATCAAC 284


RESULT 6
AAZ11236
ID   AAZ11236 standard; cDNA; 1593 BP.
XX
AC   AAZ11236;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Human pre-proheparanase coding sequence.
XX
KW   Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
KW   inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
KW   heparin degradation; anticoagulant neutralisation; asthma; CNS disease;
KW   inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
KW   tumour growth; fibroproliferative disorder; neurodegenerative disease;
KW   therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1593
FT                   /*tag=  a
FT                   /product= pre-proheparanase
XX
PN   WO9943830-A2.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999;   99WO-US01489.
XX
PR   26-MAR-1998;   98US-0079401.
PR   24-FEB-1998;   98US-0075706.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Fairbanks MB,  Heinrikson RL,  Mildner AM;
XX
DR   WPI; 1999-540598/45.
DR   P-PSDB; AAY34173.
XX
PT   New isolated platelet heparanase polypeptides, used to develop
PT   products for, e.g. wound healing and blocking angiogenesis
XX
PS   Claim 2; Fig 7; 57pp; English.
XX
CC   This sequence encodes the human pre-proheparanase of the invention. This
CC   sequence was isolated from human platelets. The heparanase can be used
CC   for identifying agents which alter heparanase activity. The heparanase
CC   can be used for wound healing or for blocking angiogenesis or
CC   inflammation. It can be used for treating e.g. psoriasis, diabetic
CC   retinopathy or solid tumours, or for the degradation of heparin and the
CC   neutralisation of heparin's anticoagulant properties during surgery.
CC   Inhibitors of heparanase activity can be used in the treatment of
CC   arthritis, asthma, and other inflammatory diseases, vascular restenosis,
CC   atherosclerosis, tumour growth and progression, fibroproliferative
CC   disorders, and central nervous system (CNS) and neurodegenerative
CC   diseases. The products can also be used for detection and diagnosis. The
CC   purified heparanase, both recombinantly produced human heparanase and
CC   heparanase isolated from human platelet activity, allows for the
CC   convenient selection of compounds having anti-heparanase activity,
CC   i.e. inhibitors of heparanase activity, by measuring inhibition of
CC   heparanase activity. Inhibition of heparanase activity can be measured by
CC   blocking heparanase-mediated release of radioactive fragments from in
CC   vivo radiolabelled (HSPG)/heparin.
XX
SQ   Sequence 1593 BP; 426 A; 370 C; 369 G; 428 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 20;  Length 1593;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        271 TTCGATCCCAAGAAGGAATCAAC 293


RESULT 7
ABZ22816
ID   ABZ22816 standard; cDNA; 1669 BP.
XX
AC   ABZ22816;
XX
DT   02-APR-2003  (first entry)
XX
DE   Human heparanase encoding cDNA SEQ ID NO:17.
XX
KW   Human; heparanase; phosphorothioate; antisense oligonucleotide;
KW   cytostatic; gene therapy; tumour; enzyme; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1638
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO2003004705-A1.
XX
PD   16-JAN-2003.
XX
PF   01-JUL-2002; 2002WO-US20636.
XX
PR   05-JUL-2001; 2001US-0899440.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Stein C;
XX
DR   WPI; 2003-201558/19.
DR   P-PSDB; ABP56822.
XX
PT   New oligonucleotide having a sequence complementary to a sequence of
PT   ribonucleic acid encoding a heparanase, useful for preparing a
PT   composition for treating tumor -
XX
PS   Disclosure; Fig 3; 48pp; English.
XX
CC   The present invention describes an oligonucleotide having a sequence
CC   complementary to a sequence of ribonucleic acid encoding a heparanase.
CC   The oligonucleotide hybridises with the ribonucleic acid under conditions
CC   of high stringency and has a sequence comprising 10-40 bp. The
CC   internucleoside linkages of the oligonucleotide comprise at least one
CC   phosphorothioate linkage. Hybridisation of the oligonucleotide to the
CC   ribonucleic acid inhibits expression of the heparanase, where inhibition
CC   of heparanase means at least a 50% reduction in the quality of
CC   heparanase. Also described: (1) a method of inhibiting expression of a
```

CC heparanase in a cell; (2) a composition comprising the above
CC oligonucleotide in an amount effective to inhibit the expression of
CC heparanase in the cell and a carrier; and (3) a method of treating a
CC tumour in a subject comprises administering to the subject an amount of
CC the above oligonucleotide effective to inhibit expression of a heparanase
CC in the subject. Heparanase antisense oligonucleotides have cytostatic
CC activity, can be used in gene therapy, and can be used for preparing a
CC composition for treating tumours. The present sequence encodes human
CC heparanase, which is given in the exemplification of the present
CC invention.
XX
SQ Sequence 1669 BP; 445 A; 396 C; 388 G; 440 T; 0 other;

Query Match 100.0%; Score 23; DB 25; Length 1669;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCGATCCCAAGAAGGAATCAAC 23
||||||||||||||||||||||
Db 316 TTCGATCCCAAGAAGGAATCAAC 338

RESULT 8
AAX37259
ID AAX37259 standard; DNA; 1713 BP.
XX
AC AAX37259;
XX
DT 21-JUL-1999 (first entry)
XX
DE Human heparanase enzyme encoding DNA.
XX
KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG; ss.
XX
OS Homo sapiens.
XX
PN WO9921975-A1.
XX
PD 06-MAY-1999.
XX
PF 28-OCT-1998; 98WO-AU00898.
XX
PR 09-DEC-1997; 97AU-0000812.
PR 28-OCT-1997; 97AU-0000062.
XX
PA (AUSU) UNIV AUSTRALIAN NAT.
XX
PI Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;
XX
DR WPI; 1999-312956/26.
DR P-PSDB; AAY17082.
XX
PT Polynucleotides encoding mammalian endoglucuronidases, especially
PT heparanases, useful to promote wound healing
XX
PS Claim 3; Page 69-73; 112pp; English.
XX
CC The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG. The
CC present sequence represents a DNA encoding human heparanase.
XX
SQ Sequence 1713 BP; 460 A; 404 C; 406 G; 443 T; 0 other;

Query Match 100.0%; Score 23; DB 20; Length 1713;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCGATCCCAAGAAGGAATCAAC 23
||||||||||||||||||||||
Db 355 TTCGATCCCAAGAAGGAATCAAC 377

RESULT 9
AAX35648
ID AAX35648 standard; cDNA; 1721 BP.
XX
AC AAX35648;
XX
DT 09-JUL-1999 (first entry)
XX
DE cDNA encoding a human heparanase protein.
XX
KW Heparanase; hpa; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW ss.
XX
OS Homo sapiens.
XX
PN WO9911798-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US17954.
XX
PR 02-JUL-1998; 98US-0109386.
PR 02-SEP-1997; 97US-0922170.
XX
PA (FRIE/) FRIEDMAN M M.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Feinstein E, Pecker I, Vlodavsky I;
XX
DR WPI; 1999-302255/25.
DR P-PSDB; AAY02345.
XX
PT New human polynucleotide useful for treating angiogenesis,
PT restenosis, and inflammation
XX
PS Claim 4; Fig 1; 63pp; English.
XX
CC The specification describes a polypeptide having heparanase (hpa)
CC activity. The recombinant protein is used as a modulator of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections
CC or disintegration of neurodegenerative plaques. Heparanase may be
CC useful for conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC and renal failure in biopsy specimens, plasma samples, and body fluids.
CC The present sequence encodes human heparanase.
XX
SQ Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

Query Match 100.0%; Score 23; DB 20; Length 1721;

```
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


RESULT 10
AAA75051
ID   AAA75051 standard; cDNA; 1721 BP.
XX
AC   AAA75051;
XX
DT   15-JAN-2001  (first entry)
XX
DE   cDNA encoding a human heparanase polypeptide.
XX
KW   Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW   heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW   wound healing; infection; burn; angiogenesis; restenosis;
KW   atherosclerosis; inflammation; neurodegenerative disease;
KW   Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1693
FT                   /*tag=  a
FT                   /product= "heparanase"
FT   stem_loop       698..724
FT                   /*tag=  b
FT                   /note= "these nucleotides are likely to be involved
FT                          in forming stem and loop structures"
XX
PN   WO200052178-A1.
XX
PD   08-SEP-2000.
XX
PF   14-FEB-2000; 2000WO-US03542.
XX
PR   01-MAR-1999;   99US-0258892.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Feinstein E;
XX
DR   WPI; 2000-579289/54.
DR   P-PSDB; AAB08849.
XX
PT   New polynucleotides encoding a polypeptide having heparanase activity,
PT   useful in wound healing and in gene therapy, particularly in treating
PT   tumour, inflammation, autoimmunity, neurodegenerative diseases  -
XX
PS   Claim 9; Fig 1; 152pp; English.
XX
CC   The present sequence encodes a human protein with heparanase catalytic
CC   activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC   particularly in treating tumour, inflammation or autoimmunity.
CC   Particularly, the polynucleotide is useful in modulating the
CC   bioavailability of heparin-binding growth factors, cellular responses
CC   to heparin-binding growth factors (e.g. bFGF) and cytokines
CC   (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
CC   cellular susceptibility to certain viral and some bacterial and protozoa
CC   infections, or disintegration of neurodegenerative plaques. The
CC   polynucleotide is also useful in wound healing (e.g. thermal, chemical
CC   or radiation burns), and in the treatment of angiogenesis, restenosis,
CC   atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC   Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC   bacterial or protozoa infections.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


RESULT 11
AAZ39195
ID   AAZ39195 standard; cDNA; 1721 BP.
XX
AC   AAZ39195;
XX
DT   02-MAR-2000  (first entry)
XX
DE   Human heparanase encoding cDNA.
XX
KW   Human; heparanase; hpa; genetic modification; expression; anticancer;
KW   angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW   anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW   heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW   metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW   neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW   micrometastasis; autoimmune lesion; kidney failure; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1694
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO9957244-A1.
XX
PD   11-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09256.
XX
PR   01-MAY-1998;   98US-0071618.
PR   02-MAR-1999;   99US-0260038.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Ben-Artzi H,  Ayal-Hershkovitz M,  Yacoby-Zeevi O,  Pecker I,  Peleg Y;
PI   Shlomi Y;
XX
DR   WPI; 2000-062144/05.
DR   P-PSDB; AAY57590.
XX
PT   Engineered cells that express recombinant heparanase, useful
PT   therapeutically, e.g. for treating angiogenesis and to screen for
PT   specific inhibitors, potential anticancer agents  -
XX
PS   Claim 2; Page 106-107; 118pp; English.
XX
CC   The present invention describes genetically modified cells (A) containing
CC   a polynucleotide (I) that encodes a polypeptide with heparanase activity,
CC   and express recombinant heparanase (II). Heparanase cleaves heparan
CC   sulphate (HS) at specific intrachain sites, resulting in release of
CC   heparin-binding growth factors, enzymes and proteins that are sequestered
CC   by HS in basement membranes, extracellular matrix or cell surfaces. It
CC   may also be implicated in tumour angiogenesis and metastases. (II) is
CC   potentially useful in wound healing and for treating angiogenesis,
CC   restenosis, atherosclerosis, inflammation, neurodegeneration, viral
CC   infection and cystic fibrosis. It can also be used to neutralise heparin
CC   (an alternative to protamine) and to screen for specific inhibitors
CC   (potentially useful for treating cancer and metastases). Antibodies
CC   raised against (II) are used for immunodetection and diagnosis of
```

```
CC   micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
CC   in large quantities, in a form that is homogeneously processed and
CC   activated/neutralised by a dedicated protease. The present sequence
CC   encodes human heparanase.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


RESULT 12
AAZ33290
ID   AAZ33290 standard; cDNA; 1721 BP.
XX
AC   AAZ33290;
XX
DT   21-FEB-2000  (first entry)
XX
DE   Human heparanase nucleotide sequence.
XX
KW   Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW   antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW   metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW   mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW   inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW   autoimmune disease; anticancer; kidney disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1694
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO9957153-A1.
XX
PD   11-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09255.
XX
PR   01-MAY-1998;   98US-0071739.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Friedman Y,  Perets T;
XX
DR   WPI; 2000-052944/04.
DR   P-PSDB; AAY52990.
XX
PT   Heparanase-specific molecular probes useful for diagnosis and
PT   treatment, e.g. of tumors, and for targeted drug delivery -
XX
PS   Example; Page 82-84; 90pp; English.
XX
CC   The present invention describes heparanase-specific molecular probes,
CC   useful for methods of detecting heparanase in situ. The probes and
CC   anti-heparanase antibodies are used to detect or quantify the expression
CC   of heparanase, for diagnosis and monitoring of diseases (especially
CC   metastasis), for treatment of heparanase-associated diseases (e.g.
CC   tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC   mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC   metastases) derived from liver, prostate, bladder, breast, ovary,
CC   cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC   disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC   syndrome, sepsis and inflammatory or autoimmune disease), for targeted
CC   drug delivery (e.g. of anticancer agents) and as research reagents.
CC   The present sequence encodes human heparanase, which is used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


RESULT 13
AAA91112
ID   AAA91112 standard; DNA; 1721 BP.
XX
AC   AAA91112;
XX
DT   20-APR-2001  (first entry)
XX
DE   Human heparanase, coding sequence fragment isolated from EST clone.
XX
KW   Heparanase; hnhpl; wound healing; angiogenesis; restenosis; Scrape;
KW   atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW   neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW   gene therapy; mouse; expressed sequence tag; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100643-A2.
XX
PD   04-JAN-2001.
XX
PF   19-JUN-2000; 2000WO-IL00358.
XX
PR   25-JUN-1999;   99US-0140801.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI   Pecker I,  Michal I,  Itzhaki H;
XX
DR   WPI; 2001-137930/14.
XX
PT   New polynucleotides and polypeptides that are distantly homologous to
PT   heparanase, useful in wound healing, as well as in gene therapy
PT   protocols for angiogenesis, restenosis, atherosclerosis, or
PT   inflammation -
XX
PS   Example 1; Page 67; 67pp; English.
XX
CC   This sequence represents a human heparanase coding sequence clone,
CC   isolated from an EST clone. The invention relates to heparanase DNA
CC   and protein sequences. The heparanase DNA and protein sequences are
CC   useful in wound healing, angiogenesis, restenosis, atherosclerosis,
CC   inflammation, pulmonary diseases, neurodegenerative diseases (such as
CC   Scrape, Alzheimer's disease, and Creutzfeldt-Jakob disease) or viral
CC   infections. The heparanase coding sequence is particularly useful in gene
CC   therapy.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 22;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394
```

```
RESULT 14
AAF93788
ID   AAF93788 standard; cDNA; 1722 BP.
XX
AC   AAF93788;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0090.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88361.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 89; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 1722 BP; 449 A; 414 C; 412 G; 447 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 22;  Length 1722;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        373 TTCGATCCCAAGAAGGAATCAAC 395
```

```
RESULT 15
AAX37260
ID   AAX37260 standard; DNA; 1723 BP.
XX
AC   AAX37260;
XX
DT   21-JUL-1999  (first entry)
XX
DE   Seq ID No: 14 of WO9921975.
XX
KW   Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW   metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW   arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW   human; HSPG; ss.
XX
OS   Homo sapiens.
XX
PN   WO9921975-A1.
XX
PD   06-MAY-1999.
XX
PF   28-OCT-1998;   98WO-AU00898.
XX
PR   09-DEC-1997;   97AU-0000812.
PR   28-OCT-1997;   97AU-0000062.
XX
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Freeman CG,  Hamdorf BJ,  Hulett MD,  Parish CR;
XX
DR   WPI; 1999-312956/26.
DR   P-PSDB; AAY17083.
XX
PT   Polynucleotides encoding mammalian endoglucuronidases, especially
PT   heparanases, useful to promote wound healing
XX
PS   Claim 11; Page 76-79; 112pp; English.
XX
CC   The invention relates to nucleic acid sequences that encode heparanase
CC   enzymes having  endoglucuronidase activity. Recombinant heparanases are
CC   capable of removing the HS side chain from heparan sulfate proteoglycan
CC   (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC   inhibit heparanase, this is useful for treatment of a physiological or
CC   medical condition associated with elevated heparanase activity, such as
CC   metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC   arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC   rat heparanases can be used to enhance wound healing, especially
CC   associated with tissue development and repair. The conditions mentioned
CC   above can be diagnosed using specific antibodies, and also using primers
CC   and probes specific for the heparanase polynucleotides. Other uses of the
CC   heparanases include sequencing sulfated molecules such as HSPG.
XX
SQ   Sequence 1723 BP; 461 A; 407 C; 412 G; 443 T; 0 other;

  Query Match             100.0%;  Score 23;  DB 20;  Length 1723;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        361 TTCGATCCCAAGAAGGAATCAAC 383


Search completed: February 16, 2004, 09:18:13
Job time : 13.2849 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:56:25 ; Search time 109.813 Seconds
(without alignments).
8568.399 Million cell updates/sec

Title: US-10-676-079-6
Perfect score: 23
Sequence: 1 ttcgatcccaagaaggaatcaac 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	100.0	23	6	AR080677	AR080677 Sequence
2	23	100.0	23	6	AR125607	AR125607 Sequence
3	23	100.0	23	6	AR287439	AR287439 Sequence
4	23	100.0	23	6	BD074425	BD074425 Polynucle
c 5	23	100.0	558	6	BD110974	BD110974 EST and e
6	23	100.0	1593	6	AR210040	AR210040 Sequence
7	23	100.0	1593	6	BD136761	BD136761 Human pla
8	23	100.0	1669	9	AF084467	AF084467 Homo sapi
9	23	100.0	1694	9	AF152376	AF152376 Homo sapi
10	23	100.0	1713	6	AR156691	AR156691 Sequence
11	23	100.0	1713	6	AX034643	AX034643 Sequence
12	23	100.0	1721	6	AR080679	AR080679 Sequence
13	23	100.0	1721	6	AR080680	AR080680 Sequence
14	23	100.0	1721	6	AR125603	AR125603 Sequence
15	23	100.0	1721	6	AR125604	AR125604 Sequence
16	23	100.0	1721	6	AR194189	AR194189 Sequence
17	23	100.0	1721	6	AR194190	AR194190 Sequence
18	23	100.0	1721	6	AR221285	AR221285 Sequence
19	23	100.0	1721	6	AR221286	AR221286 Sequence
20	23	100.0	1721	6	AR243203	AR243203 Sequence
21	23	100.0	1721	6	AR243204	AR243204 Sequence
22	23	100.0	1721	6	AR287435	AR287435 Sequence
23	23	100.0	1721	6	AR287436	AR287436 Sequence
24	23	100.0	1721	6	BD074427	BD074427 Polynucle
25	23	100.0	1721	6	BD074428	BD074428 Polynucle
26	23	100.0	1722	6	AX136167	AX136167 Sequence
27	23	100.0	1722	6	BD123536	BD123536 Secretory
28	23	100.0	1722	9	AK075400	AK075400 Homo sapi
29	23	100.0	1723	6	AR156692	AR156692 Sequence
30	23	100.0	1723	6	AX034645	AX034645 Sequence
31	23	100.0	1724	6	AX147946	AX147946 Sequence
32	23	100.0	1724	9	AF165154	AF165154 Homo sapi
33	23	100.0	1758	9	AF144325	AF144325 Homo sapi
34	23	100.0	1810	9	BC051321	BC051321 Homo sapi
35	23	100.0	1899	6	BD074430	BD074430 Polynucle
36	23	100.0	1899	6	BD074431	BD074431 Polynucle
37	23	100.0	3726	6	AR235866	AR235866 Sequence
38	23	100.0	3726	6	AX019348	AX019348 Sequence
39	23	100.0	3726	6	BD131218	BD131218 Human hep
40	23	100.0	3726	9	AF155510	AF155510 Homo sapi
c 41	23	100.0	149188	9	AC114781	AC114781 Homo sapi
42	22	95.7	474	6	AX136496	AX136496 Sequence
43	22	95.7	474	6	BD123736	BD123736 Secretory
44	18.8	81.7	191721	10	AC101883	AC101883 Mus muscu
c 45	18.8	81.7	248754	2	AC134101	AC134101 Rattus no

ALIGNMENTS

RESULT 1
AR080677
LOCUS AR080677 23 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 6 from patent US 5968822.
ACCESSION AR080677
VERSION AR080677.1 GI:10007407
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 23)
AUTHORS Pecker,I., Vlodavsky,I. and Feinstein,E.
TITLE Polynucleotide encoding a polypeptide having heparanase activity and expression of same in transduced cells
JOURNAL Patent: US 5968822-A 6 19-OCT-1999;

```
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="unknown"
BASE COUNT        9 a      6 c      4 g      4 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db          1 TTCGATCCCAAGAAGGAATCAAC 23


RESULT 2
AR125607
LOCUS       AR125607                  23 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 6 from patent US 6177545.
ACCESSION   AR125607
VERSION     AR125607.1  GI:14111669
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6177545-A 6 23-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="unknown"
BASE COUNT        9 a      6 c      4 g      4 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db          1 TTCGATCCCAAGAAGGAATCAAC 23


RESULT 3
AR287439
LOCUS       AR287439                  23 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 6 from patent US 6531129.
ACCESSION   AR287439
VERSION     AR287439.1  GI:29725133
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6531129-A 6 11-MAR-2003;
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="unknown"
BASE COUNT        9 a      6 c      4 g      4 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db          1 TTCGATCCCAAGAAGGAATCAAC 23


RESULT 4
BD074425
LOCUS       BD074425                  23 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell.
ACCESSION   BD074425
VERSION     BD074425.1  GI:22620028
KEYWORDS    JP 2001514855-A/6.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Pecker,I., Vlodavsky,I. and Elena,F.
  TITLE     Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell
  JOURNAL   Patent: JP 2001514855-A 6 18-SEP-2001;
            INSIGHT STRATEGY & MARKETING LTD, HADASIT MEDICAL RESEARCH SERVICES
            & DEVELOPMENT LTD
COMMENT     OS   Nucleic acid
            PN   JP 2001514855-A/6
            PD   18-SEP-2001
            PF   31-AUG-1998 JP 2000508806
            PR   02-SEP-1997 US    08/922170,02-JUL-1998 US    09/109386 PI
            IRIS PECKER,ISRAEL VLODAVSKY,FEINSTEIN ELENA
            PC   C12N15/09,A61K38/00,A61P9/10,A61P17/00,A61P29/00,A61P35/00, PC
              A61P37/00,
            PC   A61P43/00,C12N5/10,C12N9/24,C12Q1/68,G01N33/15,G01N33/50// PC
             A61K39/395,
            PC   A61K39/395,C12N15/00,A61K37/02,C12N5/00
            CC   Polynucleotide encoding polypeptide having
            heparanase activity
            CC            and
            CC   expression of the polypeptide in induced cell FH   Key
                Location/Qualifiers
            FT   source          1. .23
            FT                   /organism='Nucleic acid'.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT        9 a      6 c      4 g      4 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db          1 TTCGATCCCAAGAAGGAATCAAC 23


RESULT 5
BD110974/c
LOCUS       BD110974                  558 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  EST and encoded human protein.
ACCESSION   BD110974
VERSION     BD110974.1  GI:23205792
KEYWORDS    JP 2002010789-A/3051.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Edwards,J.B.D.M., Jobert,S. and Giordano,J.E.
  TITLE     EST and encoded human protein
  JOURNAL   Patent: JP 2002010789-A 3051 15-JAN-2002;
            GENSET CORP
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002010789-A/3051
```

```
            PD   15-JAN-2002
            PF   07-AUG-2000 JP 2000280989
            PR   05-AUG-1999 US    60/147499
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,SEVELIN JOBERT,JEAN EVE  PI
             GIORDANO
            PC   C12N15/09,C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19, PC
             C12N1/21,
            PC   C12N5/10,C12P21/02,C12P21/08,C12Q1/68,C12N15/00,C12N5/00, PC
            C12N15/00
            CC   EST and encoded human protein
            FH   Key             Location/Qualifiers
            FT   CDS             140. .403.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      181 a    111 c    108 g    157 t      1 others
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        173 TTCGATCCCAAGAAGGAATCAAC 151


RESULT 6
AR210040
LOCUS       AR210040                1593 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6387643.
ACCESSION   AR210040
VERSION     AR210040.1  GI:21512167
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Heinrikson,R.Leroy., Fairbanks,M.B. and Mildner,A.M.
  TITLE     Human platelet heparanase polypeptides, polynucleotide molecules
            that encode them, and methods for the identification of compounds
            that alter heparanase activity
  JOURNAL   Patent: US 6387643-A 1 14-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="unknown"
BASE COUNT      426 a    370 c    369 g    428 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 1593;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        271 TTCGATCCCAAGAAGGAATCAAC 293


RESULT 7
BD136761
LOCUS       BD136761                1593 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Human platelet heparanase polypeptide, polynucleotide molecule
            encoding the same and method of identifying compound changing
            heparanase activity.
ACCESSION   BD136761
VERSION     BD136761.1  GI:23231706
KEYWORDS    JP 2002504376-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Heinrikson,R.L., Fairbanks,M.B. and Mildner,A.M.
  TITLE     Human platelet heparanase polypeptide, polynucleotide molecule
            encoding the same and method of identifying compound changing
  JOURNAL   Patent: JP 2002504376-A 1 12-FEB-2002;
            PHARMACIA & UPJOHN CO
COMMENT     OS   Unidentified
            PN   JP 2002504376-A/1
            PD   12-FEB-2002
            PF   18-FEB-1999 JP 2000533569
            PR   24-FEB-1998 US    60/075706,26-MAR-1998 US    60/079401 PI
            ROBERT L HEINRIKSON,MICHAEL B FAIRBANKS,ANA M MILDNER PC
            C12N15/09,C07K16/40,C12N1/21,C12N5/10,C12N9/24,C12Q1/34,C12N15/ PC
             00,C12N5/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            CC   Human platelet heparanase polypeptide, polynucleotide molecule

            CC            encoding
            CC   the same and method of identifying compound changing  CC
                 heparanase activity
            FH   Key             Location/Qualifiers
            FT   source          1. .1593
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      426 a    370 c    369 g    428 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 1593;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        271 TTCGATCCCAAGAAGGAATCAAC 293


RESULT 8
AF084467
LOCUS       AF084467                1669 bp    mRNA    linear   PRI 18-OCT-2000
DEFINITION  Homo sapiens heparanase mRNA, complete cds.
ACCESSION   AF084467
VERSION     AF084467.1  GI:5870623
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1669)
  AUTHORS   Dempsey,L.A., Plummer,T.B., Coombes,S.L. and Platt,J.L.
  TITLE     Heparanase expression in invasive trophoblasts and acute vascular
            damage
  JOURNAL   Glycobiology 10 (5), 467-475 (2000)
  MEDLINE   20229546
   PUBMED   10764835
REFERENCE   2  (bases 1 to 1669)
  AUTHORS   Dempsey,L.A., Holzknecht,R.A. and Platt,J.L.
  TITLE     Identification of the cDNA encoding human heparanase
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1669)
  AUTHORS   Dempsey,L.A., Holzknecht,R.A. and Platt,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1998) Surgery, Duke University, Research Dr., Rm.
            401 MSRB, Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .1669
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="mekakaryocyte"
```

```
                     /tissue_type="placenta"
     CDS             1. .1638
                     /note="endoglycosidase; expressed in placenta and
                     platelets"
                     /codon_start=1
                     /product="heparanase"
                     /protein_id="AAD54516.1"
                     /db_xref="GI:5870624"
                     /translation="MLLRSKPALPPPLLMLLLLGPLGPLSPGALPRPAQAQQDVVDLD
                     FFTQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTD
                     FLIFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQK
                     KFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGY
                     NISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQPRRK
                     TAKMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVES
                     TRPGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGN
                     YHLVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTNTDNPRYKEG
                     DLTLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQ
                     TLPPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI"
BASE COUNT      445 a    396 c    388 g    440 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 9;  Length 1669;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        316 TTCGATCCCAAGAAGGAATCAAC 338


RESULT 9
AF152376
LOCUS       AF152376               1694 bp    mRNA    linear   PRI 28-JUL-1999
DEFINITION  Homo sapiens heparanase mRNA, complete cds.
ACCESSION   AF152376
VERSION     AF152376.1  GI:5616196
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1694)
  AUTHORS   Kussie,P.H., Hulmes,J.D., Ludwig,D.L., Patel,S., Navarro,E.C.,
            Seddon,A.P., Giorgio,N.A. and Bohlen,P.
  TITLE     Cloning and functional expression of a human heparanase gene
  JOURNAL   Biochem. Biophys. Res. Commun. 261 (1), 183-187 (1999)
  MEDLINE   99335379
   PUBMED   10405343
REFERENCE   2  (bases 1 to 1694)
  AUTHORS   Kussie,P.H., Hulmes,J.D., Ludwig,D., Patel,S., Navarro,E.C.,
            Seddon,A.P., Giorgio,N.A. and Bohlen,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-1999) Protein Chemistry, ImClone Systems Inc.,
            180 Varick Street, New York, NY 10014, USA
FEATURES             Location/Qualifiers
     source          1. .1694
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
     CDS             15. .1646
                     /codon_start=1
                     /product="heparanase"
                     /protein_id="AAD45669.1"
                     /db_xref="GI:5616197"
                     /translation="MLLRSKPALPPPLMLLLLGPLGPLSPGALPRPAQADVVDLDFF
                     TQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTDFL
                     IFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQKKF
                     KNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNI
                     SWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQPRRKTA
                     KMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTR
                     PGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYH
                     LVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTNTDNPRYKEGDL
                     TLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQTL
                     PPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI"
BASE COUNT      465 a    398 c    391 g    440 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 9;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        324 TTCGATCCCAAGAAGGAATCAAC 346


RESULT 10
AR156691
LOCUS       AR156691               1713 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 12 from patent US 6242238.
ACCESSION   AR156691
VERSION     AR156691.1  GI:15125395
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1713)
  AUTHORS   Freeman,C.Geoffrey., Hulett,M.Darren., Parish,C.Richard. and
            Hamdorf,B.James.
  TITLE     Isolated nucleic acid molecule encoding mammalian endoglucuronidase
            and uses therefor
  JOURNAL   Patent: US 6242238-A 12 05-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .1713
                     /organism="unknown"
BASE COUNT      460 a    404 c    406 g    443 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 6;  Length 1713;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        355 TTCGATCCCAAGAAGGAATCAAC 377


RESULT 11
AX034643
LOCUS       AX034643               1713 bp    DNA     linear   PAT 22-SEP-2000
DEFINITION  Sequence 12 from Patent EP1032656.
ACCESSION   AX034643
VERSION     AX034643.1  GI:10303224
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hamdorf,B.J., Freeman,C.G., Hulett,M.D. and Parish,C.R.
  TITLE     Isolated nucleic acid molecule encoding mammalian endoglucuronidase
            and uses therefor
  JOURNAL   Patent: EP 1032656-A 12 06-SEP-2000;
            UNIV AUSTRALIAN (AU)
FEATURES             Location/Qualifiers
     source          1. .1713
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             46. .1677
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC10139.1"
                     /db_xref="GI:10303225"
                     /translation="MLLRSKPALPPPLMLLLLGPLGPLSPGALPRPAQAQDVVDLDFF
```

```
                     TQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTDFL
                     IFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQKKF
                     KNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNI
                     SWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQPRRKTA
                     KMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTR
                     PGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYH
                     LVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTNTDNPRYKEGDL
                     TLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQTL
                     PPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI"
     mat_peptide     517. .1674
                     /product="unnamed"
BASE COUNT      460 a    404 c    406 g    443 t
ORIGIN

  Query Match            100.0%;  Score 23;  DB 6;  Length 1713;
  Best Local Similarity  100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        355 TTCGATCCCAAGAAGGAATCAAC 377


RESULT 12
AR080679
LOCUS       AR080679               1721 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 9 from patent US 5968822.
ACCESSION   AR080679
VERSION     AR080679.1  GI:10007409
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I. and Feinstein,E.
  TITLE     Polynucleotide encoding a polypeptide having heparanase activity
            and expression of same in transduced cells
  JOURNAL   Patent: US 5968822-A 9 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match            100.0%;  Score 23;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


RESULT 13
AR080680
LOCUS       AR080680               1721 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 11 from patent US 5968822.
ACCESSION   AR080680
VERSION     AR080680.1  GI:10007410
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I. and Feinstein,E.
  TITLE     Polynucleotide encoding a polypeptide having heparanase activity
            and expression of same in transduced cells
  JOURNAL   Patent: US 5968822-A 11 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match            100.0%;  Score 23;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


RESULT 14
AR125603
LOCUS       AR125603               1721 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 1 from patent US 6177545.
ACCESSION   AR125603
VERSION     AR125603.1  GI:14111665
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6177545-A 1 23-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match            100.0%;  Score 23;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


RESULT 15
AR125604
LOCUS       AR125604               1721 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 3 from patent US 6177545.
ACCESSION   AR125604
VERSION     AR125604.1  GI:14111666
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6177545-A 3 23-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match            100.0%;  Score 23;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 2.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        372 TTCGATCCCAAGAAGGAATCAAC 394


Search completed: February 16, 2004, 11:43:02
Job time : 109.813 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 08:49:50 ; Search time 89.3432 Seconds
(without alignments)
6256.802 Million cell updates/sec

Title: US-10-676-079-6
Perfect score: 23
Sequence: 1 ttcgatcccaagaaggaatcaac 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	100.0	881	14	CB988510	CB988510 AGENCOURT
2	23	100.0	924	13	BQ691142	BQ691142 AGENCOURT
3	23	100.0	1185	9	AL552174	AL552174 AL552174
4	23	100.0	1201	9	AL545270	AL545270 AL545270
c 5	18.8	81.7	494	29	CC066085	CC066085 CSU-K33r.
c 6	18.8	81.7	711	29	BZ527295	BZ527295 OGAEJ93TC
c 7	18.8	81.7	711	29	BZ539243	BZ539243 OGAHJ62TC
c 8	18.4	80.0	581	9	AI965004	AI965004 fc85e07.y
9	18.4	80.0	592	14	CB091033	CB091033 gy83f07.g
10	18.4	80.0	594	14	CB091179	CB091179 he89d02.g
c 11	18.2	79.1	401	9	AI624038	AI624038 ts25h10.x
c 12	18.2	79.1	483	13	BU821234	BU821234 UB20CPG09
c 13	18.2	79.1	519	13	BU833017	BU833017 T040G08 P
14	18.2	79.1	568	29	CC327702	CC327702 OGOBS65TH
15	18.2	79.1	603	28	BH549506	BH549506 BOHSC40TF
c 16	18.2	79.1	623	13	BU835590	BU835590 T075H04 P
c 17	18.2	79.1	669	13	BU837501	BU837501 T102E06 P
c 18	18.2	79.1	696	13	BU892096	BU892096 P059D01 P
c 19	18.2	79.1	717	13	BU831251	BU831251 T019D07 P
20	18.2	79.1	812	28	BH439125	BH439125 BOHRS23TF
c 21	18.2	79.1	853	29	BZ409786	BZ409786 OGACF81TC
22	18.2	79.1	890	29	BZ973851	BZ973851 PUGIW57TB
c 23	18.2	79.1	948	29	CC008795	CC008795 PUDHM49TD
24	18	78.3	645	14	CA621440	CA621440 w11n.pk00
c 25	17.8	77.4	393	9	AJ225474	AJ225474 AJ225474
26	17.8	77.4	400	29	FR0048882	AL604691 Fugu rubr
27	17.8	77.4	497	28	AQ716252	AQ716252 HS_5452_B
28	17.8	77.4	1072	29	CNS054DM	AL320611 Tetraodon
29	17.8	77.4	1078	10	BE615628	BE615628 601278645
30	17.4	75.7	420	29	CC054600	CC054600 SALK_0778
31	17.4	75.7	572	29	CC414518	CC414518 PUEGS42TD
32	17.4	75.7	597	28	AQ563293	AQ563293 HS_5334_B
33	17.4	75.7	659	14	CB449302	CB449302 703490 MA
34	17.4	75.7	701	29	AG105115	AG105115 Pan trogl
c 35	17.4	75.7	716	28	BH095882	BH095882 RPCI-24-2
c 36	17.4	75.7	720	28	BH427866	BH427866 BOGMK60TR
37	17.4	75.7	787	28	BH525146	BH525146 BOHSZ81TF
c 38	17.4	75.7	800	29	BZ285719	BZ285719 CH230-255
c 39	17.4	75.7	1071	28	BZ165314	BZ165314 CH230-462
c 40	17.2	74.8	140	29	BZ539173	BZ539173 OGAFV20TC
41	17.2	74.8	175	29	BZ712362	BZ712362 OGDAM29TM
c 42	17.2	74.8	178	29	BZ712351	BZ712351 OGDAM29TC
c 43	17.2	74.8	186	29	BZ659850	BZ659850 OGANU61TC
44	17.2	74.8	197	9	AV358527	AV358527 AV358527
c 45	17.2	74.8	239	13	BQ454119	BQ454119 sao76c06.

ALIGNMENTS

RESULT 1
CB988510
LOCUS CB988510 881 bp mRNA linear EST 01-MAY-2003
DEFINITION AGENCOURT_13905817 NIH_MGC_147 Homo sapiens cDNA clone IMAGE:30340461 5', mRNA sequence.
ACCESSION CB988510
VERSION CB988510.1 GI:30283030
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 881)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM370 row: f column: 22

```
          High quality sequence stop: 664.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30340461"
                     /tissue_type="Human Placenta"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_147"
                     /note="Organ: placenta; Vector: pBluescriptR; Site_1:
                     alI-XhoI; Site_2: BamH; Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: This is
                     a NIH_MGC library."
BASE COUNT      200 a    244 c    229 g    208 t
ORIGIN

  Query Match             100.0%;  Score 23;  DB 14;  Length 881;
  Best Local Similarity   100.0%;  Pred. No. 4.3;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        431 TTCGATCCCAAGAAGGAATCAAC 453


RESULT 2
BQ691142
LOCUS       BQ691142                924 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8343629 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250265
            5', mRNA sequence.
ACCESSION   BQ691142
VERSION     BQ691142.1  GI:21816458
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2393  row: a  column: 18
            High quality sequence stop: 710.
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6250265"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      203 a    271 c    227 g    219 t      4 others
ORIGIN

  Query Match             100.0%;  Score 23;  DB 13;  Length 924;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        444 TTCGATCCCAAGAAGGAATCAAC 466


RESULT 3
AL552174
LOCUS       AL552174                1185 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL552174 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI059YN15 5-PRIME, mRNA sequence.
ACCESSION   AL552174
VERSION     AL552174.2  GI:31273990
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1185)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12890820.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI059CG08QP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI059CG08QP1.
FEATURES             Location/Qualifiers
     source          1. .1185
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI059YN15"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      280 a    293 c    293 g    280 t      39 others
ORIGIN

  Query Match             100.0%;  Score 23;  DB 9;  Length 1185;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        438 TTCGATCCCAAGAAGGAATCAAC 460


RESULT 4
AL545270
LOCUS       AL545270                1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL545270 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI028YF04 5-PRIME, mRNA sequence.
ACCESSION   AL545270
```

```
VERSION     AL545270.2  GI:31267106
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12877751.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI028DC02QP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI028DC02QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI028YF04"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      292 a    282 c    305 g    279 t     43 others
ORIGIN

  Query Match             100.0%;  Score 23;  DB 9;  Length 1201;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              |||||||||||||||||||||||
Db        431 TTCGATCCCAAGAAGGAATCAAC 453


RESULT 5
CC066085/c
LOCUS       CC066085                494 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.41K4.T7 CSU-K33r Aedes aegypti genomic clone CSU-K33r.41K4
            , genomic survey sequence.
ACCESSION   CC066085
VERSION     CC066085.1  GI:29904591
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CSU-K33r.41K4.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.41K4"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
BASE COUNT      177 a     78 c    123 g    116 t
ORIGIN

  Query Match             81.7%;  Score 18.8;  DB 29;  Length 494;
  Best Local Similarity   90.9%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAA 22
              |||||| |||||||||| |||
Db         95 TTCGATACCAAGAAGGAAGCAA 74


RESULT 6
BZ527295/c
LOCUS       BZ527295                711 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  OGAEJ93TC ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMa0042O18,
            genomic survey sequence.
ACCESSION   BZ527295
VERSION     BZ527295.1  GI:27067857
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek
            ,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0042O18"
                     /clone_lib="ZM2_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
BASE COUNT      161 a    186 c    158 g    206 t
ORIGIN

  Query Match             81.7%;  Score 18.8;  DB 29;  Length 711;
  Best Local Similarity   90.9%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAA 22
              || |||| ||||||||||||||
Db        353 TTTGATCGCAAGAAGGAATCAA 332


RESULT 7
```

```
BZ539243/c
LOCUS       BZ539243                711 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  OGAHJ62TC ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMa0061L04,
            genomic survey sequence.
ACCESSION   BZ539243
VERSION     BZ539243.1  GI:27087679
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek
            ,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0061L04"
                     /clone_lib="ZM2_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
BASE COUNT      163 a    183 c    167 g    198 t
ORIGIN

  Query Match             81.7%;  Score 18.8;  DB 29;  Length 711;
  Best Local Similarity   90.9%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAA 22
              || |||| ||||||||||||||||
Db        396 TTTGATCGCAAGAAGGAATCAA 375


RESULT 8
AI965004/c
LOCUS       AI965004                581 bp    mRNA    linear   EST 07-JUN-2001
DEFINITION  fc85e07.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
            IMAGE:3728196 5' similar to gb:X17206 40S RIBOSOMAL PROTEIN S4
            (HUMAN);, mRNA sequence.
ACCESSION   AI965004
VERSION     AI965004.1  GI:5759641
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished
COMMENT     Other_ESTs: fc85e07.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Trace considered overall poor quality
            zebrafish identity (p-value greater than 1e-99) found to: TIGR:
            TC299 NULL
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3728196"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3']
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      134 a    154 c    166 g    127 t
ORIGIN

  Query Match             80.0%;  Score 18.4;  DB 9;  Length 581;
  Best Local Similarity   95.0%;  Pred. No. 5.4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 CGATCCCAAGAAGGAATCAA 22
              |||||||||||||||| ||||
Db        162 CGATCCCAAGAAGGAGTCAA 143


RESULT 9
CB091033
LOCUS       CB091033                592 bp    mRNA    linear   EST 27-JAN-2003
DEFINITION  gy83f07.g1 Cycad Leaf Library (NYBG) Cycas rumphii cDNA clone
            gy83f07, mRNA sequence.
ACCESSION   CB091033
VERSION     CB091033.1  GI:27915225
KEYWORDS    EST.
SOURCE      Cycas rumphii
  ORGANISM  Cycas rumphii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Cycadophyta; Cycadales; Cycadaceae; Cycas.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Brenner,E.D., Katari,M.S., Dedhia,N.N., O'Shaughnessy,A.L., Balija
```

```
          ,V., Martienssen,R.A., McCombie,R.W., Benfey,P., Coruzzi,G. and
          Stevenson,D.
  TITLE   Expressed tag sequences from Cycas ovule (NYBG)
  JOURNAL Unpublished
COMMENT   Contact: W. Richard McCombie
          Lita Annenberg Hazen Genome Sequencing Center
          Cold Spring Harbor Laboratory
          PO Box 100, Cold Spring Harbor, NY 11724, USA
          Tel: 516 367 8884
          Fax: 516 367 8874
          Email: mccombie@cshl.org
          Plate: gy83  row: f  column: 07
          Seq primer: -21M13UnivRev
          High quality sequence stop: 592.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Cycas rumphii"
                     /mol_type="mRNA"
                     /db_xref="taxon:58031"
                     /clone="gy83f07"
                     /sex="Female"
                     /clone_lib="Cycad Leaf Library (NYBG)"
                     /note="Organ: Young leaf; Vector: pBK-CMV; Site_1: Xho I;
                     Site_2: Eco RI; Date: Completed 09/01/2001. Submitted to
                     CSHL 09/05/2001. Sample: Young emergent leaves. From New
                     York Botanical Garden Conservatory accssion number 808/59
                     A (collected 03/2001). Library: Made using Stratagene's
                     ZAP Express Vector Kit. Library was size fractionated for
                     large inserts."
BASE COUNT      148 a    129 c    149 g    166 t
ORIGIN

  Query Match             80.0%;  Score 18.4;  DB 14;  Length 592;
  Best Local Similarity   95.0%;  Pred. No. 5.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GATCCCAAGAAGGAATCAAC 23
              | ||||||||||||||||||
Db        302 GCTCCCAAGAAGGAATCAAC 321


RESULT 10
CB091179
LOCUS       CB091179                 594 bp    mRNA    linear   EST 27-JAN-2003
DEFINITION  he89d02.g1 Cycad Leaf Library (NYBG) Cycas rumphii cDNA clone
          he89d02, mRNA sequence.
ACCESSION   CB091179
VERSION     CB091179.1  GI:27915371
KEYWORDS    EST.
SOURCE      Cycas rumphii
  ORGANISM  Cycas rumphii
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Cycadophyta; Cycadales; Cycadaceae; Cycas.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Brenner,E.D., Katari,M.S., Dedhia,N.N., O'Shaughnessy,A.L., Balija
          ,V., Martienssen,R.A., McCombie,R.W., Benfey,P., Coruzzi,G. and
          Stevenson,D.
  TITLE     Expressed tag sequences from Cycas ovule (NYBG)
  JOURNAL   Unpublished
COMMENT   Contact: W. Richard McCombie
          Lita Annenberg Hazen Genome Sequencing Center
          Cold Spring Harbor Laboratory
          PO Box 100, Cold Spring Harbor, NY 11724, USA
          Tel: 516 367 8884
          Fax: 516 367 8874
          Email: mccombie@cshl.org
          Plate: he89  row: d  column: 02
          Seq primer: -21M13UnivRev
          High quality sequence stop: 594.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Cycas rumphii"
                     /mol_type="mRNA"
                     /db_xref="taxon:58031"
                     /clone="he89d02"
                     /sex="Female"
                     /clone_lib="Cycad Leaf Library (NYBG)"
                     /note="Organ: Young leaf; Vector: pBK-CMV; Site_1: Xho I;
                     Site_2: Eco RI; Date: Completed 09/01/2001. Submitted to
                     CSHL 09/05/2001. Sample: Young emergent leaves. From New
                     York Botanical Garden Conservatory accssion number 808/59
                     A (collected 03/2001). Library: Made using Stratagene's
                     ZAP Express Vector Kit. Library was size fractionated for
                     large inserts."
BASE COUNT      159 a    127 c    144 g    164 t
ORIGIN

  Query Match             80.0%;  Score 18.4;  DB 14;  Length 594;
  Best Local Similarity   95.0%;  Pred. No. 5.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GATCCCAAGAAGGAATCAAC 23
              | ||||||||||||||||||
Db        389 GCTCCCAAGAAGGAATCAAC 408


RESULT 11
AI624038/c
LOCUS       AI624038                 401 bp    mRNA    linear   EST 14-DEC-1999
DEFINITION  ts25h10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2229667 3'
          similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AI624038
VERSION     AI624038.1  GI:4648969
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Life Technologies catalog #: 11548-013
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Insert Length: 835   Std Error: 0.00
          Seq primer: -40UP from Gibco
          High quality sequence stop: 400
          POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .401
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2229667"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Pan1"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT       92 a     86 c    100 g    123 t
ORIGIN

  Query Match             79.1%;  Score 18.2;  DB 9;  Length 401;
  Best Local Similarity   87.0%;  Pred. No. 6.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              || |||||| |||||||| ||||
```

```
Db        229 TTTGATCCCCAGAAGGAAGCAAC 207


RESULT 12
BU821234/c
LOCUS       BU821234                483 bp    mRNA    linear   EST 15-OCT-2002
DEFINITION  UB20CPG09 Populus tremula cambium cDNA library Populus tremula cDNA
            5 prime, mRNA sequence.
ACCESSION   BU821234
VERSION     BU821234.1  GI:23987320
KEYWORDS    EST.
SOURCE      Populus tremula
  ORGANISM  Populus tremula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Populus tremula"
                     /mol_type="mRNA"
                     /db_xref="taxon:113636"
                     /tissue_type="cambium"
                     /clone_lib="Populus tremula cambium cDNA library"
BASE COUNT      118 a    165 c     84 g    116 t
ORIGIN

  Query Match           79.1%;  Score 18.2;  DB 13;  Length 483;
  Best Local Similarity   87.0%;  Pred. No. 6.4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              || | |||||||||||  ||||||
Db        479 TTTGCTCCCAAGAAGGTATCAAC 457


RESULT 13
BU833017/c
LOCUS       BU833017                519 bp    mRNA    linear   EST 15-OCT-2002
DEFINITION  T040G08 Populus apical shoot cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU833017
VERSION     BU833017.1  GI:24013112
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="apical shoot"
                     /clone_lib="Populus apical shoot cDNA library"
BASE COUNT      123 a    127 c    125 g    144 t
ORIGIN

  Query Match           79.1%;  Score 18.2;  DB 13;  Length 519;
  Best Local Similarity   87.0%;  Pred. No. 6.5e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              || | |||||||||||  ||||||
Db         47 TTTGCTCCCAAGAAGGTATCAAC 25


RESULT 14
CC327702
LOCUS       CC327702                568 bp    DNA     linear   GSS 16-MAY-2003
DEFINITION  OGOBS65THB ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0369K09,
            genomic survey sequence.
ACCESSION   CC327702
VERSION     CC327702.1  GI:30796873
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek
            ,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0369K09"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
BASE COUNT      212 a     73 c     85 g    198 t
ORIGIN

  Query Match           79.1%;  Score 18.2;  DB 29;  Length 568;
  Best Local Similarity   87.0%;  Pred. No. 6.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              || || | ||||||||||||||||
Db        296 TTAGAACACAAGAAGGAATCAAC 318


RESULT 15
BH549506
LOCUS       BH549506                603 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOHSC40TF BOHS Brassica oleracea genomic clone BOHSC40, genomic
            survey sequence.
ACCESSION   BH549506
```

```
VERSION     BH549506.1  GI:17801286
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BOHSC40TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHSC40"
                     /clone_lib="BOHS"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      189 a    147 c     94 g    173 t
ORIGIN

  Query Match             79.1%;  Score 18.2;  DB 28;  Length 603;
  Best Local Similarity   87.0%;  Pred. No. 6.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TTCGATCCCAAGAAGGAATCAAC 23
              || ||| |||||||||||| ||||
Db        236 TTAGATTCCAAGAAGGAAGCAAC 258


Search completed: February 16, 2004, 13:41:01
Job time : 92.3432 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:56:25 ; Search time 114.587 Seconds
(without alignments)
8568.399 Million cell updates/sec

Title: US-10-676-079-7
Perfect score: 24
Sequence: 1 gtagtgatgccatgtaactgaatc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24	100.0	24	6	AR080673	AR080673 Sequence
2	24	100.0	24	6	AR080678	AR080678 Sequence
3	24	100.0	24	6	AR125608	AR125608 Sequence
4	24	100.0	24	6	AR287440	AR287440 Sequence
5	24	100.0	24	6	BD074421	BD074421 Polynucle
6	24	100.0	24	6	BD074426	BD074426 Polynucle
c 7	24	100.0	1593	6	AR210040	AR210040 Sequence
c 8	24	100.0	1593	6	BD136761	BD136761 Human pla
c 9	24	100.0	1662	4	AF281160	AF281160 Bos tauru
c 10	24	100.0	1669	9	AF084467	AF084467 Homo sapi
c 11	24	100.0	1694	9	AF152376	AF152376 Homo sapi
c 12	24	100.0	1713	6	AR156691	AR156691 Sequence
c 13	24	100.0	1713	6	AX034643	AX034643 Sequence
c 14	24	100.0	1721	6	AR080679	AR080679 Sequence
c 15	24	100.0	1721	6	AR080680	AR080680 Sequence
c 16	24	100.0	1721	6	AR125603	AR125603 Sequence
c 17	24	100.0	1721	6	AR125604	AR125604 Sequence
c 18	24	100.0	1721	6	AR194189	AR194189 Sequence
c 19	24	100.0	1721	6	AR194190	AR194190 Sequence
c 20	24	100.0	1721	6	AR221285	AR221285 Sequence
c 21	24	100.0	1721	6	AR221286	AR221286 Sequence
c 22	24	100.0	1721	6	AR243203	AR243203 Sequence
c 23	24	100.0	1721	6	AR243204	AR243204 Sequence
c 24	24	100.0	1721	6	AR287435	AR287435 Sequence
c 25	24	100.0	1721	6	AR287436	AR287436 Sequence
c 26	24	100.0	1721	6	BD074427	BD074427 Polynucle
c 27	24	100.0	1721	6	BD074428	BD074428 Polynucle
c 28	24	100.0	1722	6	AX136167	AX136167 Sequence
c 29	24	100.0	1722	6	BD123536	BD123536 Secretory
c 30	24	100.0	1722	9	AK075400	AK075400 Homo sapi
c 31	24	100.0	1723	6	AR156692	AR156692 Sequence
c 32	24	100.0	1723	6	AX034645	AX034645 Sequence
c 33	24	100.0	1724	6	AX147946	AX147946 Sequence
c 34	24	100.0	1724	9	AF165154	AF165154 Homo sapi
c 35	24	100.0	1758	9	AF144325	AF144325 Homo sapi
c 36	24	100.0	1810	9	BC051321	BC051321 Homo sapi
c 37	24	100.0	1899	6	BD074430	BD074430 Polynucle
c 38	24	100.0	1899	6	BD074431	BD074431 Polynucle
c 39	24	100.0	3726	6	AR235866	AR235866 Sequence
c 40	24	100.0	3726	6	AX019348	AX019348 Sequence
c 41	24	100.0	3726	6	BD131218	BD131218 Human hep
c 42	24	100.0	3726	9	AF155510	AF155510 Homo sapi
43	21.4	89.2	149188	9	AC114781	AC114781 Homo sapi
44	19.8	82.5	118890	2	AC138159	AC138159 Rattus no
c 45	19.8	82.5	176787	2	AC138158	AC138158 Rattus no

ALIGNMENTS

RESULT 1
AR080673
LOCUS AR080673 24 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 2 from patent US 5968822.
ACCESSION AR080673
VERSION AR080673.1 GI:10007403
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Pecker,I., Vlodavsky,I. and Feinstein,E.
TITLE Polynucleotide encoding a polypeptide having heparanase activity and expression of same in transduced cells
JOURNAL Patent: US 5968822-A 2 19-OCT-1999;

```
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        7 a      4 c      6 g      7 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 2
AR080678
LOCUS       AR080678                  24 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 7 from patent US 5968822.
ACCESSION   AR080678
VERSION     AR080678.1  GI:10007408
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Pecker,I., Vlodavsky,I. and Feinstein,E.
  TITLE     Polynucleotide encoding a polypeptide having heparanase activity
            and expression of same in transduced cells
  JOURNAL   Patent: US 5968822-A 7 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        7 a      4 c      6 g      7 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 3
AR125608
LOCUS       AR125608                  24 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 7 from patent US 6177545.
ACCESSION   AR125608
VERSION     AR125608.1  GI:14111670
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6177545-A 7 23-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        7 a      4 c      6 g      7 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 4
AR287440
LOCUS       AR287440                  24 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 7 from patent US 6531129.
ACCESSION   AR287440
VERSION     AR287440.1  GI:29725134
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6531129-A 7 11-MAR-2003;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        7 a      4 c      6 g      7 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 5
BD074421
LOCUS       BD074421                  24 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell.
ACCESSION   BD074421
VERSION     BD074421.1  GI:22620024
KEYWORDS    JP 2001514855-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Pecker,I., Vlodavsky,I. and Elena,F.
  TITLE     Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell
  JOURNAL   Patent: JP 2001514855-A 2 18-SEP-2001;
            INSIGHT STRATEGY & MARKETING LTD, HADASIT MEDICAL RESEARCH SERVICES
            & DEVELOPMENT LTD
COMMENT     OS   Nucleic acid
            PN   JP 2001514855-A/2
            PD   18-SEP-2001
            PF   31-AUG-1998 JP 2000508806
            PR   02-SEP-1997 US    08/922170,02-JUL-1998 US    09/109386 PI
            IRIS PECKER,ISRAEL VLODAVSKY,FEINSTEIN ELENA
            PC   C12N15/09,A61K38/00,A61P9/10,A61P17/00,A61P29/00,A61P35/00, PC
              A61P37/00,
            PC   A61P43/00,C12N5/10,C12N9/24,C12Q1/68,G01N33/15,G01N33/50// PC
             A61K39/395,
            PC   A61K39/395,C12N15/00,A61K37/02,C12N5/00
            CC   Polynucleotide encoding polypeptide having
            heparanase activity
            CC          and
            CC   expression of the polypeptide in induced cell FH   Key
                Location/Qualifiers
            FT   source          1. .24
            FT                   /organism='Nucleic acid'.
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
```

```
BASE COUNT       7 a      4 c      6 g      7 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.69;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 6
BD074426
LOCUS       BD074426                 24 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell.
ACCESSION   BD074426
VERSION     BD074426.1  GI:22620029
KEYWORDS    JP 2001514855-A/7.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Pecker,I., Vlodavsky,I. and Elena,F.
  TITLE     Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell
  JOURNAL   Patent: JP 2001514855-A 7 18-SEP-2001;
            INSIGHT STRATEGY & MARKETING LTD, HADASIT MEDICAL RESEARCH SERVICES
            & DEVELOPMENT LTD
COMMENT     OS   Nucleic acid
            PN   JP 2001514855-A/7
            PD   18-SEP-2001
            PF   31-AUG-1998 JP 2000508806
            PR   02-SEP-1997 US    08/922170,02-JUL-1998 US    09/109386 PI
            IRIS PECKER,ISRAEL VLODAVSKY,FEINSTEIN ELENA
            PC   C12N15/09,A61K38/00,A61P9/10,A61P17/00,A61P29/00,A61P35/00, PC
              A61P37/00,
            PC   A61P43/00,C12N5/10,C12N9/24,C12Q1/68,G01N33/15,G01N33/50// PC
             A61K39/395,
            PC   A61K39/395,C12N15/00,A61K37/02,C12N5/00
            CC   Polynucleotide encoding polypeptide having
            heparanase activity
            CC            and
            CC   expression of the polypeptide in induced cell FH   Key
                Location/Qualifiers
            FT   source          1. .24
            FT                   /organism='Nucleic acid'.
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT       7 a      4 c      6 g      7 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.69;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 7
AR210040/c
LOCUS       AR210040                 1593 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6387643.
ACCESSION   AR210040
VERSION     AR210040.1  GI:21512167
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Heinrikson,R.Leroy., Fairbanks,M.B. and Mildner,A.M.
  TITLE     Human platelet heparanase polypeptides, polynucleotide molecules
            that encode them, and methods for the identification of compounds
            that alter heparanase activity
  JOURNAL   Patent: US 6387643-A 1 14-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="unknown"
BASE COUNT      426 a    370 c    369 g    428 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 6;  Length 1593;
  Best Local Similarity  100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        855 GTAGTGATGCCATGTAACTGAATC 832


RESULT 8
BD136761/c
LOCUS       BD136761                 1593 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Human platelet heparanase polypeptide, polynucleotide molecule
            encoding the same and method of identifying compound changing
            heparanase activity.
ACCESSION   BD136761
VERSION     BD136761.1  GI:23231706
KEYWORDS    JP 2002504376-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Heinrikson,R.L., Fairbanks,M.B. and Mildner,A.M.
  TITLE     Human platelet heparanase polypeptide, polynucleotide molecule
            encoding the same and method of identifying compound changing
  JOURNAL   Patent: JP 2002504376-A 1 12-FEB-2002;
            PHARMACIA & UPJOHN CO
COMMENT     OS   Unidentified
            PN   JP 2002504376-A/1
            PD   12-FEB-2002
            PF   18-FEB-1999 JP 2000533569
            PR   24-FEB-1998 US    60/075706,26-MAR-1998 US    60/079401 PI
            ROBERT L HEINRIKSON,MICHAEL B FAIRBANKS,ANA M MILDNER PC
            C12N15/09,C07K16/40,C12N1/21,C12N5/10,C12N9/24,C12Q1/34,C12N15/ PC
             00,C12N5/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            CC   Human platelet heparanase polypeptide, polynucleotide molecule

            CC            encoding
            CC   the same and method of identifying compound changing  CC
                heparanase activity
            FH   Key             Location/Qualifiers
            FT   source          1. .1593
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      426 a    370 c    369 g    428 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 6;  Length 1593;
  Best Local Similarity  100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
```

```
Db        855 GTAGTGATGCCATGTAACTGAATC 832


RESULT 9
AF281160/c
LOCUS       AF281160               1662 bp    mRNA    linear   MAM 12-APR-2001
DEFINITION  Bos taurus heparanase mRNA, complete cds.
ACCESSION   AF281160
VERSION     AF281160.2  GI:13606094
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1662)
  AUTHORS   Kizaki,K., Nakano,H., Takahashi,T., Imai,K. and Hashizume,K.
  TITLE     Expression of Heparanase mRNA in Bovine Placenta During Gestation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1662)
  AUTHORS   Kizaki,K., Nakano,H., Takahashi,T., Imai,K. and Hashizume,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2000) Laboratory of Reproductive Endocrinology,
            National Institute of Animal Industry, Tsukuba-norindanchi, PO Box
            5, Tsukuba, Ibaraki 305-0901, Japan
REFERENCE   3  (bases 1 to 1662)
  AUTHORS   Kizaki,K., Nakano,H., Takahashi,T., Imai,K. and Hashizume,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2001) Laboratory of Reproductive Endocrinology,
            National Institute of Animal Industry, Tsukuba-norindanchi, PO Box
            5, Tsukuba, Ibaraki 305-0901, Japan
  REMARK    Sequence update by submitter
COMMENT     On Apr 12, 2001 this sequence version replaced gi:9408602.
FEATURES             Location/Qualifiers
     source          1. .1662
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="placenta"
     CDS             25. .1662
                     /note="endoglycosidase"
                     /codon_start=1
                     /product="heparanase"
                     /protein_id="AAF87301.2"
                     /db_xref="GI:13606095"
                     /translation="MLACRKPGLRPPLLLLLPLLGPLGPCSPGTPAAAAPADDAAELE
                     FFTERPLHLVSPAFLSFTIDANLATDPRFFTFLGSSKLRTLARGLAPAYLRFGGNKGD
                     FLIFDPKKEPAFEERSYWLSQSNQDICKSGSIPSDVEEKLRLEWPFQEQVLLREQYQK
                     KFTNSTYSRSSVDMLYTFASCSGLNLIFGVNALLRTTDMHWDSSNAQLLLDYCSSKNY
                     NISWELGNEPNSFQRKAGIFINGRQLGEDFIEFRKLLGKSAFKNAKLYGPDIGQPRRN
                     TVKMLKSFLKAGGEVIDSVTWHHYYVNGRIATKEDFLNPDILDTFISSVQKTLRIVEK
                     IRPLKKVWLGETSSAFGGGAPFLSNTFAAGFMWLDKLGLSARMGIEVVMRQVLFGAGN
                     YHLVDGNFEPLPDYWLSLLFKKLVGNKVLMASVKGPDRSKFRVYLHCTNTKHPRYKEG
                     DLTLYALNLHNVTKHLELPHHLFNKQVDKYLIKPSGTDGLLSKSVQLNGQILKMVDEQ
                     TLPALTEKPLHPGSSLGMPPFSYGFFVIRNAKVAACI"
BASE COUNT      435 a    404 c    402 g    421 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 4;  Length 1662;
  Best Local Similarity  100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        924 GTAGTGATGCCATGTAACTGAATC 901


RESULT 10
AF084467/c
LOCUS       AF084467               1669 bp    mRNA    linear   PRI 18-OCT-2000
DEFINITION  Homo sapiens heparanase mRNA, complete cds.
ACCESSION   AF084467
VERSION     AF084467.1  GI:5870623
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1669)
  AUTHORS   Dempsey,L.A., Plummer,T.B., Coombes,S.L. and Platt,J.L.
  TITLE     Heparanase expression in invasive trophoblasts and acute vascular
            damage
  JOURNAL   Glycobiology 10 (5), 467-475 (2000)
  MEDLINE   20229546
   PUBMED   10764835
REFERENCE   2  (bases 1 to 1669)
  AUTHORS   Dempsey,L.A., Holzknecht,R.A. and Platt,J.L.
  TITLE     Identification of the cDNA encoding human heparanase
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1669)
  AUTHORS   Dempsey,L.A., Holzknecht,R.A. and Platt,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1998) Surgery, Duke University, Research Dr., Rm.
            401 MSRB, Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .1669
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="mekakaryocyte"
                     /tissue_type="placenta"
     CDS             1. .1638
                     /note="endoglycosidase; expressed in placenta and
                     platelets"
                     /codon_start=1
                     /product="heparanase"
                     /protein_id="AAD54516.1"
                     /db_xref="GI:5870624"
                     /translation="MLLRSKPALPPPLLMLLLLGPLGPLSPGALPRPAQAQQDVVDLD
                     FFTQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTD
                     FLIFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQK
                     KFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGY
                     NISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQPRRK
                     TAKMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVES
                     TRPGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGN
                     YHLVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTNTDNPRYKEG
                     DLTLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQ
                     TLPPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI"
BASE COUNT      445 a    396 c    388 g    440 t
ORIGIN

  Query Match           100.0%;  Score 24;  DB 9;  Length 1669;
  Best Local Similarity  100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        900 GTAGTGATGCCATGTAACTGAATC 877


RESULT 11
AF152376/c
LOCUS       AF152376               1694 bp    mRNA    linear   PRI 28-JUL-1999
DEFINITION  Homo sapiens heparanase mRNA, complete cds.
ACCESSION   AF152376
VERSION     AF152376.1  GI:5616196
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1694)
  AUTHORS   Kussie,P.H., Hulmes,J.D., Ludwig,D.L., Patel,S., Navarro,E.C.,
            Seddon,A.P., Giorgio,N.A. and Bohlen,P.
  TITLE     Cloning and functional expression of a human heparanase gene
  JOURNAL   Biochem. Biophys. Res. Commun. 261 (1), 183-187 (1999)
```

```
  MEDLINE   99335379
   PUBMED   10405343
REFERENCE   2  (bases 1 to 1694)
  AUTHORS   Kussie,P.H., Hulmes,J.D., Ludwig,D., Patel,S., Navarro,E.C.,
            Seddon,A.P., Giorgio,N.A. and Bohlen,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-1999) Protein Chemistry, ImClone Systems Inc.,
            180 Varick Street, New York, NY 10014, USA
FEATURES             Location/Qualifiers
     source          1. .1694
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
     CDS             15. .1646
                     /codon_start=1
                     /product="heparanase"
                     /protein_id="AAD45669.1"
                     /db_xref="GI:5616197"
                     /translation="MLLRSKPALPPPLMLLLLGPLGPLSPGALPRPAQAQDVVDLDFF
                     TQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTDFL
                     IFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQKKF
                     KNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNI
                     SWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQPRRKTA
                     KMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTR
                     PGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYH
                     LVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTNTDNPRYKEGDL
                     TLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQTL
                     PPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI"
BASE COUNT      465 a    398 c    391 g    440 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 9;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        908 GTAGTGATGCCATGTAACTGAATC 885


RESULT 12
AR156691/c
LOCUS       AR156691                1713 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 12 from patent US 6242238.
ACCESSION   AR156691
VERSION     AR156691.1  GI:15125395
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1713)
  AUTHORS   Freeman,C.Geoffrey., Hulett,M.Darren., Parish,C.Richard. and
            Hamdorf,B.James.
  TITLE     Isolated nucleic acid molecule encoding mammalian endoglucuronidase
            and uses therefor
  JOURNAL   Patent: US 6242238-A 12 05-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .1713
                     /organism="unknown"
BASE COUNT      460 a    404 c    406 g    443 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 1713;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        939 GTAGTGATGCCATGTAACTGAATC 916


RESULT 13
AX034643/c
LOCUS       AX034643                1713 bp    DNA     linear   PAT 22-SEP-2000
DEFINITION  Sequence 12 from Patent EP1032656.
ACCESSION   AX034643
VERSION     AX034643.1  GI:10303224
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hamdorf,B.J., Freeman,C.G., Hulett,M.D. and Parish,C.R.
  TITLE     Isolated nucleic acid molecule encoding mammalian endoglucuronidase
            and uses therefor
  JOURNAL   Patent: EP 1032656-A 12 06-SEP-2000;
            UNIV AUSTRALIAN (AU)
FEATURES             Location/Qualifiers
     source          1. .1713
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             46. .1677
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC10139.1"
                     /db_xref="GI:10303225"
                     /translation="MLLRSKPALPPPLMLLLLGPLGPLSPGALPRPAQAQDVVDLDFF
                     TQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTDFL
                     IFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQKKF
                     KNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNI
                     SWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQPRRKTA
                     KMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTR
                     PGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYH
                     LVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTNTDNPRYKEGDL
                     TLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQTL
                     PPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI"
     mat_peptide     517. .1674
                     /product="unnamed"
BASE COUNT      460 a    404 c    406 g    443 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 1713;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        939 GTAGTGATGCCATGTAACTGAATC 916


RESULT 14
AR080679/c
LOCUS       AR080679                1721 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 9 from patent US 5968822.
ACCESSION   AR080679
VERSION     AR080679.1  GI:10007409
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I. and Feinstein,E.
  TITLE     Polynucleotide encoding a polypeptide having heparanase activity
            and expression of same in transduced cells
  JOURNAL   Patent: US 5968822-A 9 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
```

```
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        956 GTAGTGATGCCATGTAACTGAATC 933


RESULT 15
AR080680/c
LOCUS       AR080680               1721 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 11 from patent US 5968822.
ACCESSION   AR080680
VERSION     AR080680.1  GI:10007410
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I. and Feinstein,E.
  TITLE     Polynucleotide encoding a polypeptide having heparanase activity
            and expression of same in transduced cells
  JOURNAL   Patent: US 5968822-A 11 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        956 GTAGTGATGCCATGTAACTGAATC 933


Search completed: February 16, 2004, 11:43:03
Job time : 115.587 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:55:00 ; Search time 11.7756 Seconds
(without alignments)
5501.769 Million cell updates/sec

Title: US-10-676-079-7
Perfect score: 24
Sequence: 1 gtagtgatgccatgtaactgaatc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	100.0	24	20	AAX35643	PCR primer used to
	2	24	100.0	24	20	AAX35647	PCR primer used to
	3	24	100.0	24	21	AAA75045	PCR primer HPL229
	4	24	100.0	24	21	AAA75050	PCR primer HPL229
	5	24	100.0	24	21	AAA75067	PCR primer Hpl 229
	6	24	100.0	24	21	AAZ33294	Human heparanase P
c	7	24	100.0	1584	24	ABL40753	Chicken signal pep
c	8	24	100.0	1593	20	AAZ11236	Human pre-prohepar
c	9	24	100.0	1669	25	ABZ22816	Human heparanase e
c	10	24	100.0	1713	20	AAX37259	Human heparanase e
c	11	24	100.0	1721	20	AAX35648	cDNA encoding a hu
c	12	24	100.0	1721	21	AAA75051	cDNA encoding a hu
c	13	24	100.0	1721	21	AAZ39195	Human heparanase e
c	14	24	100.0	1721	21	AAZ33290	Human heparanase n
c	15	24	100.0	1721	22	AAA91112	Human heparanase,
c	16	24	100.0	1722	22	AAF93788	Human cDNA encodin
c	17	24	100.0	1723	20	AAX37260	Seq ID No: 14 of W
c	18	24	100.0	1724	22	AAH20940	Human heparanase i
c	19	24	100.0	1899	20	AAX35650	cDNA encoding a hu
c	20	24	100.0	1899	21	AAA75053	cDNA encoding a hu
c	21	24	100.0	3726	20	AAX86671	cDNA encoding a hu
c	22	21.4	89.2	44848	21	AAA75080	Nucleotide sequenc
c	23	18.2	75.8	1380	20	AAX37261	Murine heparanase
c	24	18.2	75.8	2396	21	AAA75081	cDNA encoding a mu
c	25	18.2	75.8	2396	22	AAA91113	Mouse heparanase c
	26	17.6	73.3	242	20	AAZ33700	Tobacco plant resi
c	27	17.6	73.3	2275	22	AAH01252	Enterococcus faeca
c	28	17.2	71.7	123	24	ABV87579	Human colon cancer
c	29	17.2	71.7	397	22	AAF64922	Novel human polynu
c	30	17.2	71.7	1511	24	AAD29205	Human heparanase-2
c	31	17.2	71.7	1674	24	AAD29203	Human heparanase-2
c	32	17.2	71.7	1685	24	AAD29204	Human heparanase-2
c	33	17.2	71.7	1847	24	AAD29202	Human heparanase-2
c	34	16.8	70.0	400	23	ABV53985	Human prostate exp
c	35	16.8	70.0	474	22	AAS26020	Human cDNA encodin
c	36	16.8	70.0	474	25	ABX73361	Human novel polynu
c	37	16.8	70.0	1996	22	AAS26656	Human genomic DNA
c	38	16.8	70.0	1996	25	ABX74005	Human novel polynu
c	39	16.8	70.0	2126	22	AAS26657	Human genomic DNA
c	40	16.8	70.0	2126	22	AAS26659	Human genomic DNA
c	41	16.8	70.0	2126	25	ABX74006	Human novel polynu
c	42	16.8	70.0	2126	25	ABX74008	Human novel polynu
c	43	16.8	70.0	110000	22	AAF84800	Nucleotide sequenc
	44	16.8	70.0	218336	25	ABQ76678	Androgen receptor
c	45	16.6	69.2	188	21	AAC15008	Human secreted pro

ALIGNMENTS

RESULT 1
AAX35643
ID AAX35643 standard; DNA; 24 BP.
XX
AC AAX35643;
XX
DT 09-JUL-1999 (first entry)
XX
DE PCR primer used to amplify human hp3 cDNA.
XX
KW Heparanase; hp; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherscleros is; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW PCR primer; ss.
XX
OS Synthetic.
XX
PN WO9911798-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US17954.
XX
PR 02-JUL-1998; 98US-0109386.
PR 02-SEP-1997; 97US-0922170.
XX
PA (FRIE/) FRIEDMAN M M.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Feinstein E, Pecker I, Vlodavsky I;
XX
DR WPI; 1999-302255/25.
XX
PT New human polynucleotide useful for treating angiogenesis,
PT restenosis, and inflammation
XX
PS Example 1; Page 22; 63pp; English.
XX
CC The specification describes a polypeptide having heparanase (hp)
CC activity. The recombinant protein is used as a modulator of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections
CC or disintegration of neurodegenerative plaques. Heparanase may be
CC useful for conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC and renal failure in biopsy specimens, plasma samples, and body fluids.
CC PCR primers AAX35642-43 were used to amplify hp3 cDNA, in the course of
CC the invention.
XX
SQ Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTAGTGATGCCATGTAACTGAATC 24
||||||||||||||||||||||||
Db 1 GTAGTGATGCCATGTAACTGAATC 24

RESULT 2
AAX35647
ID AAX35647 standard; DNA; 24 BP.
XX
AC AAX35647;
XX
DT 09-JUL-1999 (first entry)
XX
DE PCR primer used to amplify human hp3 cDNA.
XX
KW Heparanase; hp; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW PCR primer; ss.
XX
OS Synthetic.
XX
PN WO9911798-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US17954.
XX
PR 02-JUL-1998; 98US-0109386.
PR 02-SEP-1997; 97US-0922170.
XX
PA (FRIE/) FRIEDMAN M M.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Feinstein E, Pecker I, Vlodavsky I;
XX
DR WPI; 1999-302255/25.
XX
PT New human polynucleotide useful for treating angiogenesis,
PT restenosis, and inflammation
XX
PS Example 1; Page 23; 63pp; English.
XX
CC The specification describes a polypeptide having heparanase (hp)
CC activity. The recombinant protein is used as a modulator of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections
CC or disintegration of neurodegenerative plaques. Heparanase may be
CC useful for conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC and renal failure in biopsy specimens, plasma samples, and body fluids.
CC PCR primers AAX35646-47 were used to amplify hp3 cDNA, in the course of
CC the invention.
XX
SQ Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTAGTGATGCCATGTAACTGAATC 24
||||||||||||||||||||||||
Db 1 GTAGTGATGCCATGTAACTGAATC 24

RESULT 3
AAA75045
ID AAA75045 standard; DNA; 24 BP.
XX
AC AAA75045;
XX
DT 15-JAN-2001 (first entry)
XX
DE PCR primer HPL229 used to amplify human cDNA encoding heparanase.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US03542.
XX
PR 01-MAR-1999; 99US-0258892.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
DR WPI; 2000-579289/54.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumour, inflammation, autoimmunity, neurodegenerative diseases -
XX
PS Disclosure; Page 44; 152pp; English.
XX
CC The present PCR primer was used to amplify a human cDNA sequence,

CC which encoded a protein with heparanase catalytic activity. The
CC heparanase (hpa) polynucleotide is useful in gene therapy, particularly
CC in treating tumour, inflammation or autoimmunity. Particularly, the
CC polynucleotide is useful in modulating the bioavailability of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors (e.g. bFGF) and cytokines (e.g. interleukin (IL)-8),
CC cell interaction with plasma lipoproteins, cellular susceptibility to
CC certain viral and some bacterial and protozoa infections, or
CC disintegration of neurodegenerative plaques. The polynucleotide is
CC also useful in wound healing (e.g. thermal, chemical or radiation burns),
CC and in the treatment of angiogenesis, restenosis, atherosclerosis,
CC inflammation, neurodegenerative diseases (Gerstmann-Straussler Syndrome
CC or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa
CC infections.
XX
SQ Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 24; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTAGTGATGCCATGTAACTGAATC 24
||||||||||||||||||||||||
Db 1 GTAGTGATGCCATGTAACTGAATC 24

RESULT 4
AAA75050
ID AAA75050 standard; DNA; 24 BP.
XX
AC AAA75050;
XX
DT 15-JAN-2001 (first entry)
XX
DE PCR primer HPL229 used to amplify human cDNA encoding heparanase.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US03542.
XX
PR 01-MAR-1999; 99US-0258892.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
DR WPI; 2000-579289/54.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumour, inflammation, autoimmunity, neurodegenerative diseases -
XX
PS Disclosure; Page 44; 152pp; English.
XX
CC The present PCR primer was used to amplify a human cDNA sequence,
CC which encoded a protein with heparanase catalytic activity. The
CC heparanase (hpa) polynucleotide is useful in gene therapy, particularly
CC in treating tumour, inflammation or autoimmunity. Particularly, the
CC polynucleotide is useful in modulating the bioavailability of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors (e.g. bFGF) and cytokines (e.g. interleukin (IL)-8),
CC cell interaction with plasma lipoproteins, cellular susceptibility to
CC certain viral and some bacterial and protozoa infections, or
CC disintegration of neurodegenerative plaques. The polynucleotide is
CC also useful in wound healing (e.g. thermal, chemical or radiation burns),
CC and in the treatment of angiogenesis, restenosis, atherosclerosis,
CC inflammation, neurodegenerative diseases (Gerstmann-Straussler Syndrome
CC or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa
CC infections.
XX
SQ Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 24; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTAGTGATGCCATGTAACTGAATC 24
||||||||||||||||||||||||
Db 1 GTAGTGATGCCATGTAACTGAATC 24

RESULT 5
AAA75067
ID AAA75067 standard; DNA; 24 BP.
XX
AC AAA75067;
XX
DT 15-JAN-2001 (first entry)
XX
DE PCR primer Hpl 229 used to amplify human cDNA encoding heparanase.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US03542.
XX
PR 01-MAR-1999; 99US-0258892.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
DR WPI; 2000-579289/54.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumour, inflammation, autoimmunity, neurodegenerative diseases -
XX
PS Disclosure; Page 45; 152pp; English.
XX
CC The present PCR primer was used to amplify a human cDNA sequence,
CC which encoded a protein with heparanase catalytic activity. The
CC heparanase (hpa) polynucleotide is useful in gene therapy, particularly
CC in treating tumour, inflammation or autoimmunity. Particularly, the
CC polynucleotide is useful in modulating the bioavailability of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors (e.g. bFGF) and cytokines (e.g. interleukin (IL)-8),
CC cell interaction with plasma lipoproteins, cellular susceptibility to
CC certain viral and some bacterial and protozoa infections, or
CC disintegration of neurodegenerative plaques. The polynucleotide is
CC also useful in wound healing (e.g. thermal, chemical or radiation burns),
CC and in the treatment of angiogenesis, restenosis, atherosclerosis,
CC inflammation, neurodegenerative diseases (Gerstmann-Straussler Syndrome

```
CC   or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa
CC   infections.
XX
SQ   Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

  Query Match           100.0%;  Score 24;  DB 21;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.012;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 6
AAZ33294
ID   AAZ33294 standard; DNA; 24 BP.
XX
AC   AAZ33294;
XX
DT   21-FEB-2000  (first entry)
XX
DE   Human heparanase PCR primer Hpl-229 SEQ ID NO:7.
XX
KW   Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW   antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW   metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW   mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW   inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW   autoimmune disease; anticancer; kidney disease; PCR primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9957153-A1.
XX
PD   11-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09255.
XX
PR   01-MAY-1998;   98US-0071739.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Friedman Y,  Perets T;
XX
DR   WPI; 2000-052944/04.
XX
PT   Heparanase-specific molecular probes useful for diagnosis and
PT   treatment, e.g. of tumors, and for targeted drug delivery  -
XX
PS   Example; Page 30; 90pp; English.
XX
CC   The present invention describes heparanase-specific molecular probes,
CC   useful for methods of detecting heparanase in situ. The probes and
CC   anti-heparanase antibodies are used to detect or quantify the expression
CC   of heparanase, for diagnosis and monitoring of diseases (especially
CC   metastasis), for treatment of heparanase-associated diseases (e.g.
CC   tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC   mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC   metastases) derived from liver, prostate, bladder, breast, ovary,
CC   cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC   disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC   syndrome, sepsis and inflammatory or autoimmune disease), for targeted
CC   drug delivery (e.g. of anticancer agents) and as research reagents.
CC   The present sequence represents a PCR primer for human heparanase, which
CC   is used in an example from the present invention.
XX
SQ   Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

  Query Match           100.0%;  Score 24;  DB 21;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.012;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db          1 GTAGTGATGCCATGTAACTGAATC 24


RESULT 7
ABL40753/c
ID   ABL40753 standard; cDNA; 1584 BP.
XX
AC   ABL40753;
XX
DT   03-JUL-2002  (first entry)
XX
DE   Chicken signal peptide/human heparanase chimeric cDNA.
XX
KW   Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
KW   anti-protozoan; neuroprotective; heparin; chicken; human; chimeric; ss.
XX
OS   Synthetic.
OS   Gallus gallus.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1584
FT                   /*tag=  a
FT                   /product= "chimeric chicken-human heparanase"
FT   sig_peptide     1..57
FT                   /*tag=  b
FT                   /note= "chicken heparanase signal peptide"
FT   mat_peptide     58..1581
FT                   /*tag=  c
FT                   /note= "human mature heparanase"
XX
PN   US2002034810-A1.
XX
PD   21-MAR-2002.
XX
PF   16-AUG-2001; 2001US-0930218.
XX
PR   20-SEP-2000; 2000US-0666390.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI   Goldshmidt O,  Pecker I,  Vlodavsky I,  Michal I,  Zcharia E;
XX
DR   WPI; 2002-338926/37.
DR   P-PSDB; ABB07815.
XX
PT   Nucleic acid encoding avian and reptile heparanase polypeptide is
PT   useful to treat various heparin-related disorders and the signal
PT   peptide is useful in production of membrane-targeted or secreted
PT   recombinant proteins  -
XX
PS   Disclosure; Page 24-25; 39pp; English.
XX
CC   The invention relates to an isolated avian and reptile nucleic acid,
CC   encoding a polypeptide with heparanase catalytic activity. The signal
CC   peptide of the nucleic acid can be used to express membrane-associated or
CC   secreted proteins in heterologous expression systems. The encoded
CC   polypeptides can be used to prevent tumour angiogenesis, metastasis and
CC   invasion, and to intervene with pathologies associated with impaired
CC   heparin-binding growth factors, cellular responses to heparin-binding
CC   growth factors and cytokines, cell interaction with plasma lipoproteins,
CC   cellular susceptibility to viral, protozoa and bacterial infections or
CC   disintegration of neurodegenerative plaques. The present sequence
CC   represents a chicken signal peptide/human heparanase chimeric cDNA
CC   sequence.
XX
SQ   Sequence 1584 BP; 424 A; 361 C; 373 G; 426 T; 0 other;
```

```
  Query Match             100.0%;  Score 24;  DB 24;  Length 1584;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        846 GTAGTGATGCCATGTAACTGAATC 823


RESULT 8
AAZ11236/c
ID   AAZ11236 standard; cDNA; 1593 BP.
XX
AC   AAZ11236;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Human pre-proheparanase coding sequence.
XX
KW   Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
KW   inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
KW   heparin degradation; anticoagulant neutralisation; asthma; CNS disease;
KW   inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
KW   tumour growth; fibroproliferative disorder; neurodegenerative disease;
KW   therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1593
FT                   /*tag=  a
FT                   /product= pre-proheparanase
XX
PN   WO9943830-A2.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999;   99WO-US01489.
XX
PR   26-MAR-1998;   98US-0079401.
PR   24-FEB-1998;   98US-0075706.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Fairbanks MB,  Heinrikson RL,  Mildner AM;
XX
DR   WPI; 1999-540598/45.
DR   P-PSDB; AAY34173.
XX
PT   New isolated platelet heparanase polypeptides, used to develop
PT   products for, e.g. wound healing and blocking angiogenesis
XX
PS   Claim 2; Fig 7; 57pp; English.
XX
CC   This sequence encodes the human pre-proheparanase of the invention. This
CC   sequence was isolated from human platelets. The heparanase can be used
CC   for identifying agents which alter heparanase activity. The heparanase
CC   can be used for wound healing or for blocking angiogenesis or
CC   inflammation. It can be used for treating e.g. psoriasis, diabetic
CC   retinopathy or solid tumours, or for the degradation of heparin and the
CC   neutralisation of heparin's anticoagulant properties during surgery.
CC   Inhibitors of heparanase activity can be used in the treatment of
CC   arthritis, asthma, and other inflammatory diseases, vascular restenosis,
CC   atherosclerosis, tumour growth and progression, fibroproliferative
CC   disorders, and central nervous system (CNS) and neurodegenerative
CC   diseases. The products can also be used for detection and diagnosis. The
CC   purified heparanase, both recombinantly produced human heparanase and
CC   heparanase isolated from human platelet activity, allows for the
CC   convenient selection of compounds having anti-heparanase activity,
CC   i.e. inhibitors of heparanase activity, by measuring inhibition of
CC   heparanase activity. Inhibition of heparanase activity can be measured by
CC   blocking heparanase-mediated release of radioactive fragments from in
CC   vivo radiolabelled (HSPG)/heparin.
XX
SQ   Sequence 1593 BP; 426 A; 370 C; 369 G; 428 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 20;  Length 1593;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        855 GTAGTGATGCCATGTAACTGAATC 832


RESULT 9
ABZ22816/c
ID   ABZ22816 standard; cDNA; 1669 BP.
XX
AC   ABZ22816;
XX
DT   02-APR-2003  (first entry)
XX
DE   Human heparanase encoding cDNA SEQ ID NO:17.
XX
KW   Human; heparanase; phosphorothioate; antisense oligonucleotide;
KW   cytostatic; gene therapy; tumour; enzyme; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1638
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO2003004705-A1.
XX
PD   16-JAN-2003.
XX
PF   01-JUL-2002; 2002WO-US20636.
XX
PR   05-JUL-2001; 2001US-0899440.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Stein C;
XX
DR   WPI; 2003-201558/19.
DR   P-PSDB; ABP56822.
XX
PT   New oligonucleotide having a sequence complementary to a sequence of
PT   ribonucleic acid encoding a heparanase, useful for preparing a
PT   composition for treating tumor -
XX
PS   Disclosure; Fig 3; 48pp; English.
XX
CC   The present invention describes an oligonucleotide having a sequence
CC   complementary to a sequence of ribonucleic acid encoding a heparanase.
CC   The oligonucleotide hybridises with the ribonucleic acid under conditions
CC   of high stringency and has a sequence comprising 10-40 bp. The
CC   internucleoside linkages of the oligonucleotide comprise at least one
CC   phosphorothioate linkage. Hybridisation of the oligonucleotide to the
CC   ribonucleic acid inhibits expression of the heparanase, where inhibition
CC   of heparanase means at least a 50% reduction in the quality of
CC   heparanase. Also described: (1) a method of inhibiting expression of a
CC   heparanase in a cell; (2) a composition comprising the above
CC   oligonucleotide in an amount effective to inhibit the expression of
CC   heparanase in the cell and a carrier; and (3) a method of treating a
CC   tumour in a subject comprises administering to the subject an amount of
CC   the above oligonucleotide effective to inhibit expression of a heparanase
CC   in the subject. Heparanase antisense oligonucleotides have cytostatic
CC   activity, can be used in gene therapy, and can be used for preparing a
CC   composition for treating tumours. The present sequence encodes human
CC   heparanase, which is given in the exemplification of the present
CC   invention.
XX
```

```
SQ   Sequence 1669 BP; 445 A; 396 C; 388 G; 440 T; 0 other;

  Query Match           100.0%;  Score 24;  DB 25;  Length 1669;
  Best Local Similarity  100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        900 GTAGTGATGCCATGTAACTGAATC 877


RESULT 10
AAX37259/c
ID   AAX37259 standard; DNA; 1713 BP.
XX
AC   AAX37259;
XX
DT   21-JUL-1999  (first entry)
XX
DE   Human heparanase enzyme encoding DNA.
XX
KW   Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW   metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW   arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW   human; HSPG; ss.
XX
OS   Homo sapiens.
XX
PN   WO9921975-A1.
XX
PD   06-MAY-1999.
XX
PF   28-OCT-1998;   98WO-AU00898.
XX
PR   09-DEC-1997;   97AU-0000812.
PR   28-OCT-1997;   97AU-0000062.
XX
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Freeman CG,  Hamdorf BJ,  Hulett MD,  Parish CR;
XX
DR   WPI; 1999-312956/26.
DR   P-PSDB; AAY17082.
XX
PT   Polynucleotides encoding mammalian endoglucuronidases, especially
PT   heparanases, useful to promote wound healing
XX
PS   Claim 3; Page 69-73; 112pp; English.
XX
CC   The invention relates to nucleic acid sequences that encode heparanase
CC   enzymes having  endoglucuronidase activity. Recombinant heparanases are
CC   capable of removing the HS side chain from heparan sulfate proteoglycan
CC   (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC   inhibit heparanase, this is useful for treatment of a physiological or
CC   medical condition associated with elevated heparanase activity, such as
CC   metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC   arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC   rat heparanases can be used to enhance wound healing, especially
CC   associated with tissue development and repair. The conditions mentioned
CC   above can be diagnosed using specific antibodies, and also using primers
CC   and probes specific for the heparanase polynucleotides. Other uses of the
CC   heparanases include sequencing sulfated molecules such as HSPG. The
CC   present sequence represents a DNA encoding human heparanase.
XX
SQ   Sequence 1713 BP; 460 A; 404 C; 406 G; 443 T; 0 other;

  Query Match           100.0%;  Score 24;  DB 20;  Length 1713;
  Best Local Similarity  100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        939 GTAGTGATGCCATGTAACTGAATC 916


RESULT 11
AAX35648/c
ID   AAX35648 standard; cDNA; 1721 BP.
XX
AC   AAX35648;
XX
DT   09-JUL-1999  (first entry)
XX
DE   cDNA encoding a human heparanase protein.
XX
KW   Heparanase; hpa; modulator; heparin-binding growth factor;
KW   cellular response; cytokine; cell interaction; plasma lipoprotein;
KW   cellular susceptibility; infection; disintegration;
KW   neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW   athersclerosis; inflammation; neurodegenerative disease; neutralise;
KW   plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO9911798-A1.
XX
PD   11-MAR-1999.
XX
PF   31-AUG-1998;   98WO-US17954.
XX
PR   02-JUL-1998;   98US-0109386.
PR   02-SEP-1997;   97US-0922170.
XX
PA   (FRIE/) FRIEDMAN M M.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI   Feinstein E,  Pecker I,  Vlodavsky I;
XX
DR   WPI; 1999-302255/25.
DR   P-PSDB; AAY02345.
XX
PT   New human polynucleotide useful for treating angiogenesis,
PT   restenosis, and inflammation
XX
PS   Claim 4; Fig 1; 63pp; English.
XX
CC   The specification describes a polypeptide having heparanase (hpa)
CC   activity. The recombinant protein is used as a modulator of
CC   heparin-binding growth factors, cellular responses to heparin-binding
CC   growth factors and cytokines, cell interaction with plasma lipoproteins,
CC   cellular susceptibility to viral, protozoal and bacterial infections
CC   or disintegration of neurodegenerative plaques. Heparanase may be
CC   useful for conditions such as wound healing, angiogenesis, restenosis,
CC   athersclerosis, inflammation, neurodegenerative diseases, and viral
CC   infections. Mammalian heparanase can be used to neutralize plasma
CC   heparin, and anti-heparanase antibodies may be applied for
CC   immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC   and renal failure in biopsy specimens, plasma samples, and body fluids.
CC   The present sequence encodes human heparanase.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match           100.0%;  Score 24;  DB 20;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        956 GTAGTGATGCCATGTAACTGAATC 933


RESULT 12
AAA75051/c
ID   AAA75051 standard; cDNA; 1721 BP.
```

```
XX
AC   AAA75051;
XX
DT   15-JAN-2001  (first entry)
XX
DE   cDNA encoding a human heparanase polypeptide.
XX
KW   Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW   heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW   wound healing; infection; burn; angiogenesis; restenosis;
KW   atherosclerosis; inflammation; neurodegenerative disease;
KW   Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1693
FT                   /*tag=  a
FT                   /product= "heparanase"
FT   stem_loop       698..724
FT                   /*tag=  b
FT                   /note= "these nucleotides are likely to be involved
FT                           in forming stem and loop structures"
XX
PN   WO200052178-A1.
XX
PD   08-SEP-2000.
XX
PF   14-FEB-2000; 2000WO-US03542.
XX
PR   01-MAR-1999;   99US-0258892.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Feinstein E;
XX
DR   WPI; 2000-579289/54.
DR   P-PSDB; AAB08849.
XX
PT   New polynucleotides encoding a polypeptide having heparanase activity,
PT   useful in wound healing and in gene therapy, particularly in treating
PT   tumour, inflammation, autoimmunity, neurodegenerative diseases  -
XX
PS   Claim 9; Fig 1; 152pp; English.
XX
CC   The present sequence encodes a human protein with heparanase catalytic
CC   activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC   particularly in treating tumour, inflammation or autoimmunity.
CC   Particularly, the polynucleotide is useful in modulating the
CC   bioavailability of heparin-binding growth factors, cellular responses
CC   to heparin-binding growth factors (e.g. bFGF) and cytokines
CC   (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
CC   cellular susceptibility to certain viral and some bacterial and protozoa
CC   infections, or disintegration of neurodegenerative plaques. The
CC   polynucleotide is also useful in wound healing (e.g. thermal, chemical
CC   or radiation burns), and in the treatment of angiogenesis, restenosis,
CC   atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC   Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC   bacterial or protozoa infections.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        956 GTAGTGATGCCATGTAACTGAATC 933
```

```
RESULT 13
AAZ39195/c
ID   AAZ39195 standard; cDNA; 1721 BP.
XX
AC   AAZ39195;
XX
DT   02-MAR-2000  (first entry)
XX
DE   Human heparanase encoding cDNA.
XX
KW   Human; heparanase; hpa; genetic modification; expression; anticancer;
KW   angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW   anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW   heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW   metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW   neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW   micrometastasis; autoimmune lesion; kidney failure; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1694
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO9957244-A1.
XX
PD   11-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09256.
XX
PR   01-MAY-1998;   98US-0071618.
PR   02-MAR-1999;   99US-0260038.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Ben-Artzi H,  Ayal-Hershkovitz M,  Yacoby-Zeevi O,  Pecker I,  Peleg Y;
PI   Shlomi Y;
XX
DR   WPI; 2000-062144/05.
DR   P-PSDB; AAY57590.
XX
PT   Engineered cells that express recombinant heparanase, useful
PT   therapeutically, e.g. for treating angiogenesis and to screen for
PT   specific inhibitors, potential anticancer agents  -
XX
PS   Claim 2; Page 106-107; 118pp; English.
XX
CC   The present invention describes genetically modified cells (A) containing
CC   a polynucleotide (I) that encodes a polypeptide with heparanase activity,
CC   and express recombinant heparanase (II). Heparanase cleaves heparan
CC   sulphate (HS) at specific intrachain sites, resulting in release of
CC   heparin-binding growth factors, enzymes and proteins that are sequestered
CC   by HS in basement membranes, extracellular matrix or cell surfaces. It
CC   may also be implicated in tumour angiogenesis and metastases. (II) is
CC   potentially useful in wound healing and for treating angiogenesis,
CC   restenosis, atherosclerosis, inflammation, neurodegeneration, viral
CC   infection and cystic fibrosis. It can also be used to neutralise heparin
CC   (an alternative to protamine) and to screen for specific inhibitors
CC   (potentially useful for treating cancer and metastases). Antibodies
CC   raised against (II) are used for immunodetection and diagnosis of
CC   micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
CC   in large quantities, in a form that is homogeneously processed and
CC   activated/neutralised by a dedicated protease. The present sequence
CC   encodes human heparanase.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              |||||||||||||||||||||||||
Db        956 GTAGTGATGCCATGTAACTGAATC 933
```

RESULT 14
AAZ33290/c
ID AAZ33290 standard; cDNA; 1721 BP.
XX
AC AAZ33290;
XX
DT 21-FEB-2000 (first entry)
XX
DE Human heparanase nucleotide sequence.
XX
KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW autoimmune disease; anticancer; kidney disease; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   63..1694
FT   /*tag= a
FT   /product= "heparanase"
XX
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09255.
XX
PR 01-MAY-1998; 98US-0071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
DR P-PSDB; AAY52990.
XX
PT Heparanase-specific molecular probes useful for diagnosis and
PT treatment, e.g. of tumors, and for targeted drug delivery -
XX
PS Example; Page 82-84; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and
CC anti-heparanase antibodies are used to detect or quantify the expression
CC of heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary,
CC cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC syndrome, sepsis and inflammatory or autoimmune disease), for targeted
CC drug delivery (e.g. of anticancer agents) and as research reagents.
CC The present sequence encodes human heparanase, which is used in the
CC exemplification of the present invention.
XX
SQ Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

```
  Query Match             100.0%;  Score 24;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              |||||||||||||||||||||||||
Db        956 GTAGTGATGCCATGTAACTGAATC 933
```

RESULT 15
AAA91112/c
ID AAA91112 standard; DNA; 1721 BP.
XX
AC AAA91112;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase, coding sequence fragment isolated from EST clone.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; mouse; expressed sequence tag; ds.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL00358.
XX
PR 25-JUN-1999; 99US-0140801.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI; 2001-137930/14.
XX
PT New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX
PS Example 1; Page 67; 67pp; English.
XX
CC This sequence represents a human heparanase coding sequence clone,
CC isolated from an EST clone. The invention relates to heparanase DNA
CC and protein sequences. The heparanase DNA and protein sequences are
CC useful in wound healing, angiogenesis, restenosis, atherosclerosis,
CC inflammation, pulmonary diseases, neurodegenerative diseases (such as
CC Scrape, Alzheimer's disease, and Creutzfeldt-Jakob disease) or viral
CC infections. The heparanase coding sequence is particularly useful in gene
CC therapy.
XX
SQ Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

```
  Query Match             100.0%;  Score 24;  DB 22;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              |||||||||||||||||||||||||
Db        956 GTAGTGATGCCATGTAACTGAATC 933
```

Search completed: February 16, 2004, 09:18:14
Job time : 12.7756 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 08:49:50 ; Search time 93.2277 Seconds
(without alignments)
6256.802 Million cell updates/sec

Title: US-10-676-079-7
Perfect score: 24
Sequence: 1 gtagtgatgccatgtaactgaatc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	24	100.0	370	14	N41349	N41349 yw70a03.r1
c	2	24	100.0	587	14	N45367	N45367 yw97a02.r1
c	3	24	100.0	907	13	BQ438834	BQ438834 AGENCOURT
	4	24	100.0	1083	13	BX398409	BX398409 BX398409
	5	24	100.0	1156	9	AL552151	AL552151 AL552151
	6	24	100.0	1200	9	AL545232	AL545232 AL545232
c	7	18.8	78.3	824	28	BZ114057	BZ114057 CH230-385
	8	18.4	76.7	493	28	AQ260481	AQ260481 CITBI-E1-
c	9	18.4	76.7	642	10	BE565766	BE565766 601338571
c	10	18.4	76.7	686	10	BB181417	BB181417 BB181417
c	11	18.4	76.7	2733	11	AK039061	AK039061 Mus muscu
	12	18.2	75.8	401	28	AQ775821	AQ775821 HS_2006_A
	13	18.2	75.8	431	28	AQ363749	AQ363749 nbxb0059J
	14	18.2	75.8	473	10	BG238460	BG238460 sab51a01.
	15	18.2	75.8	540	28	B52785	B52785 CIT-HSP-200
c	16	18.2	75.8	631	13	BU352277	BU352277 603526037
c	17	18.2	75.8	716	13	BU005587	BU005587 QGG8I20.y
c	18	18.2	75.8	732	10	BE565192	BE565192 601342769
c	19	18.2	75.8	755	29	CC412604	CC412604 PUEJN06TD
c	20	18.2	75.8	757	12	BG777686	BG777686 602665307
c	21	18.2	75.8	768	14	CB970530	CB970530 CAB10003_
c	22	18.2	75.8	813	29	BZ771989	BZ771989 mct83e06_
c	23	18.2	75.8	1962	11	AK087283	AK087283 Mus muscu
c	24	18.2	75.8	2173	11	AK040471	AK040471 Mus muscu
c	25	18	75.0	114	29	AL937708	AL937708 Arabidops
	26	17.8	74.2	84	28	AZ939449	AZ939449 2M0198K20
	27	17.8	74.2	246	10	BF084900	BF084900 MR0-FT017
c	28	17.8	74.2	262	14	CB062055	CB062055 4012122 B
	29	17.8	74.2	543	29	BZ280332	BZ280332 CH230-344
c	30	17.8	74.2	565	13	BQ465842	BQ465842 HU04N13r
	31	17.8	74.2	649	10	BB218197	BB218197 BB218197
c	32	17.8	74.2	657	10	BF207911	BF207911 601862428
c	33	17.8	74.2	687	10	BF029562	BF029562 601765976
	34	17.8	74.2	692	12	BI250835	BI250835 602993460
c	35	17.8	74.2	727	10	BE567253	BE567253 601340854
c	36	17.8	74.2	748	10	BG529572	BG529572 602557984
c	37	17.8	74.2	789	10	BF211534	BF211534 601812219
c	38	17.8	74.2	829	10	BF028342	BF028342 601765135
c	39	17.8	74.2	843	10	BF382630	BF382630 601816826
c	40	17.8	74.2	851	10	BG615642	BG615642 602642992
c	41	17.8	74.2	857	10	BE865600	BE865600 601677947
c	42	17.8	74.2	877	10	BE568413	BE568413 601341993
c	43	17.8	74.2	878	10	BG532594	BG532594 602562161
c	44	17.8	74.2	900	10	BG615175	BG615175 602643702
c	45	17.8	74.2	903	10	BG615047	BG615047 602644333

ALIGNMENTS

RESULT 1
N41349/c
LOCUS N41349 370 bp mRNA linear EST 24-JAN-1996
DEFINITION yw70a03.r1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:257548 5', mRNA sequence.
ACCESSION N41349
VERSION N41349.1 GI:1165380
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 370)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 279
Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 279.

FEATURES Location/Qualifiers
source 1. .370
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3887158"
/db_xref="taxon:9606"
/clone="IMAGE:257548"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 113 a 64 c 83 g 103 t 7 others
ORIGIN

Query Match 100.0%; Score 24; DB 14; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        229 GTAGTGATGCCATGTAACTGAATC 206
```

RESULT 2
N45367/c
LOCUS N45367 587 bp mRNA linear EST 13-FEB-1996
DEFINITION yw97a02.r1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:260138 5', mRNA sequence.
ACCESSION N45367
VERSION N45367.1 GI:1186533
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 587)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 369.

FEATURES Location/Qualifiers
source 1. .587
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3889844"
/db_xref="taxon:9606"
/clone="IMAGE:260138"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 158 a 112 c 153 g 156 t 8 others
ORIGIN

Query Match 100.0%; Score 24; DB 14; Length 587;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        173 GTAGTGATGCCATGTAACTGAATC 150
```

RESULT 3
BQ438834/c
LOCUS BQ438834 907 bp mRNA linear EST 24-MAY-2002
DEFINITION AGENCOURT_7761619 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6017952 5', mRNA sequence.
ACCESSION BQ438834
VERSION BQ438834.1 GI:21177910
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 907)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM13218 row: b column: 01
High quality sequence stop: 616.

FEATURES Location/Qualifiers
source 1. .907
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6017952"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."

BASE COUNT 260 a 176 c 226 g 242 t 3 others
ORIGIN

Query Match 100.0%; Score 24; DB 13; Length 907;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
```

```
Db        576 GTAGTGATGCCATGTAACTGAATC 553


RESULT 4
BX398409
LOCUS       BX398409               1083 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX398409 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI058YI24 3-PRIME, mRNA sequence.
ACCESSION   BX398409
VERSION     BX398409.1  GI:30617572
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1083)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI058BE12NP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI058BE12NP1.
FEATURES             Location/Qualifiers
     source          1. .1083
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI058YI24"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      273 a    264 c    206 g    258 t     82 others
ORIGIN

  Query Match             100.0%;  Score 24;  DB 13;  Length 1083;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        762 GTAGTGATGCCATGTAACTGAATC 785


RESULT 5
AL552151
LOCUS       AL552151               1156 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL552151 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI059YN15 3-PRIME, mRNA sequence.
ACCESSION   AL552151
VERSION     AL552151.2  GI:31273967
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12890775.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI059CG08NP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI059CG08NP1.
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI059YN15"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      305 a    251 c    233 g    323 t     44 others
ORIGIN

  Query Match             100.0%;  Score 24;  DB 9;  Length 1156;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              ||||||||||||||||||||||||
Db        732 GTAGTGATGCCATGTAACTGAATC 755


RESULT 6
AL545232
LOCUS       AL545232               1200 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL545232 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI028YF04 3-PRIME, mRNA sequence.
ACCESSION   AL545232
VERSION     AL545232.2  GI:31267068
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12877713.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI028DC02NP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI028DC02NP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI028YF04"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
```

primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 300 a 249 c 249 g 332 t 70 others
ORIGIN

Query Match 100.0%; Score 24; DB 9; Length 1200;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTAGTGATGCCATGTAACTGAATC 24
||||||||||||||||||||||||
Db 703 GTAGTGATGCCATGTAACTGAATC 726

RESULT 7
BZ114057/c
LOCUS BZ114057 824 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-385I18.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-385I18, genomic survey sequence.
ACCESSION BZ114057
VERSION BZ114057.1 GI:23755004
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 824)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 385 row: I column: 18
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .824
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-385I18"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"
BASE COUNT 293 a 133 c 163 g 235 t
ORIGIN

Query Match 78.3%; Score 18.8; DB 28; Length 824;
Best Local Similarity 90.9%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTAGTGATGCCATGTAACTGAA 22
|||||||| |||||||| ||||
Db 817 GTAGTGATACCATGTAATTGAA 796

RESULT 8
AQ260481
LOCUS AQ260481 493 bp DNA linear GSS 24-OCT-1998
DEFINITION CITBI-E1-2504G20.TR CITBI-E1 Homo sapiens genomic clone 2504G20, genomic survey sequence.
ACCESSION AQ260481
VERSION AQ260481.1 GI:3787005
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .493
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2504G20"
/sex="male"
/cell_type="sperm"
/clone_lib="CITBI-E1"
/note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI; CalTech Human BAC Library D"
BASE COUNT 180 a 80 c 95 g 138 t
ORIGIN

Query Match 76.7%; Score 18.4; DB 28; Length 493;
Best Local Similarity 95.0%; Pred. No. 7.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 TGATGCCATGTAACTGAATC 24
|||||||||||||||||| |
Db 322 TGATGCCATGTAACTGAAAC 341

RESULT 9
BE565766/c
LOCUS BE565766 642 bp mRNA linear EST 15-AUG-2000
DEFINITION 601338571F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3680848 5', mRNA sequence.
ACCESSION BE565766
VERSION BE565766.1 GI:9809486
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

```
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM361  row: n  column: 17
            High quality sequence stop: 524.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3680848"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_53"
                     /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T).   Average
                     insert size 1.55 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      202 a    130 c    141 g    169 t
ORIGIN

  Query Match             76.7%;  Score 18.4;  DB 10;  Length 642;
  Best Local Similarity   95.0%;  Pred. No. 8.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GTGATGCCATGTAACTGAAT 23
              ||||| ||||||||||||||
Db        597 GTGATACCATGTAACTGAAT 578


RESULT 10
BB181417/c
LOCUS       BB181417                686 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  BB181417 RIKEN full-length enriched, adult male hypothalamus Mus
            musculus cDNA clone A230091D07 3', mRNA sequence.
ACCESSION   BB181417
VERSION     BB181417.2  GI:16270073
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     On Jun 29, 2000 this sequence version replaced gi:8840500.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="A230091D07"
                     /sex="male"
                     /tissue_type="hypothalamus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male
                     hypothalamus"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT      236 a    119 c    160 g    171 t
ORIGIN

  Query Match             76.7%;  Score 18.4;  DB 10;  Length 686;
  Best Local Similarity   95.0%;  Pred. No. 8.3e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 TAGTGATGCCATGTAACTGA 21
              |||||||| ||||||||||
Db         33 TAGTGATGTCATGTAACTGA 14


RESULT 11
AK039061/c
LOCUS       AK039061                2733 bp    mRNA    linear   HTC 05-DEC-2002
```

DEFINITION Mus musculus adult male hypothalamus cDNA, RIKEN full-length enriched library, clone:A230091D07 product:adenomatosis polyposis coli, full insert sequence.
ACCESSION AK039061
VERSION AK039061.1 GI:26086918
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2733)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1. .2733
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A230091D07"
/db_xref="taxon:10090"
/clone="A230091D07"
/sex="male"
/tissue_type="hypothalamus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature 1. .2733
/note="adenomatosis polyposis coli (MGD|MGI:88039, GB|NM_007462, evidence: BLASTN, 100%, match=3169)"
BASE COUNT 857 a 520 c 513 g 843 t
ORIGIN

Query Match 76.7%; Score 18.4; DB 11; Length 2733;
Best Local Similarity 95.0%; Pred. No. 1.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TAGTGATGCCATGTAACTGA 21
||||||| ||||||||||
Db 2076 TAGTGATGTCATGTAACTGA 2057

RESULT 12
AQ775821
LOCUS AQ775821 401 bp DNA linear GSS 29-JUL-1999
DEFINITION HS_2006_A2_H02_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2006 Col=4 Row=O, genomic survey sequence.
ACCESSION AQ775821
VERSION AQ775821.1 GI:5655549
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 401)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L

```
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2006  row: O  column: 4
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 401.
FEATURES              Location/Qualifiers
     source           1. .401
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="Plate=2006 Col=4 Row=O"
                      /sex="male"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      149 a     63 c     50 g    139 t
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 28;  Length 401;
  Best Local Similarity   87.0%;  Pred. No. 8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAAT 23
              | ||||||| ||||||| ||||||
Db         74 GAAGTGATTCCATGTATCTGAAT 96


RESULT 13
AQ363749
LOCUS       AQ363749                431 bp    DNA     linear   GSS 16-DEC-1999
DEFINITION  nbxb0059J12f CUGI Rice BAC Library Oryza sativa (japonica
            cultivar-group) genomic clone nbxb0059J12f, genomic survey
            sequence.
ACCESSION   AQ363749
VERSION     AQ363749.2  GI:6583615
KEYWORDS    GSS.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished
COMMENT     On Dec 15, 1999 this sequence version replaced gi:4213404.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 231.
FEATURES              Location/Qualifiers
     source           1. .431
                      /organism="Oryza sativa (japonica cultivar-group)"
                      /mol_type="genomic DNA"
                      /strain="Japonica"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:39947"
                      /clone="nbxb0059J12f"
                      /tissue_type="Leaf"
                      /lab_host="E. coli DH10B"
                      /clone_lib="CUGI Rice BAC Library"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII; Rice is one of two most popular grains in the
                      world. Half of the world  population especially those
                      inhabiting highly populated areas of the  humid tropics
                      and subtropics, rely on rice as their primary source of
                      carbohydrate. Monocotyledonous rice is a diploid plant
                      (2n=24) with a  haploid genome equivalent of 431 Mbp
                      (Arumuganathan and Earle, 1991).  The relatively small
                      genome of rice, three times larger than that of
                      Arabidopsis, makes it suitable for genomic studies. In
                      order to  facilitate positional cloning, physical mapping
                      and genome sequencing  of rice, we have constructed a BAC
                      library from Oryza sativa, Nipponbare  variety. The
                      library contains 36,864 clones with an average insert size
                      of 128.5 Kb providing 10.9 haploid genome equivalents. The
                      deep  coverage allows the isolation a particular sequence
                      with a probability  of 99.9 %. Two high density filters,
                      each containing 18,432 clones  (doubly spotted), represent
                      the whole library for colony screening."
BASE COUNT      134 a     92 c     96 g    108 t      1 others
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 28;  Length 431;
  Best Local Similarity   83.3%;  Pred. No. 8.2e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAATC 24
              |||  |||||| |||| |||||||
Db        396 GTATNGATGCCTTGTAGCTGAATC 419


RESULT 14
BG238460
LOCUS       BG238460                 473 bp    mRNA    linear   EST 28-NOV-2001
DEFINITION  sab51a01.y1 Gm-c1043 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1043-2690 5' similar to TR:O64486 O64486 F20D22.2 PROTEIN.
            ;, mRNA sequence.
ACCESSION   BG238460
VERSION     BG238460.1  GI:12773533
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com.
FEATURES              Location/Qualifiers
     source           1. .473
                      /organism="Glycine max"
                      /mol_type="mRNA"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1043-2690"
                      /tissue_type="Hypocotyl and Plumule, germinating seeds"
```

```
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1043"
                     /note="Vector: pT7T3Pac (Pharmacia); Site_1: EcoRI;
                     Site_2: NotI; This cDNA library was constructed from mRNA
                     isolated from hypocotyl and plumule tissues of seeds
                     germinated for three days of the cultivar Williams.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting  of a poly(dT) sequence with a NotI restriction
                     site. EcoRI adapters were  ligated to the blunt-ended cDNA
                     fragments followed by digestion with EcoRI  and NotI. The
                     cDNA fragments were directionally cloned into the
                     EcoRI-NotI restriction site of the pT7T3-Pac vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (Gibco BRL). This library was constructed by Dr.
                     Randy Shoemaker."
BASE COUNT      138 a     73 c     98 g    164 t
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 10;  Length 473;
  Best Local Similarity   87.0%;  Pred. No. 8.6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 TAGTGATGCCATGTAACTGAATC 24
              || |||||||||| || ||||||
Db        442 TATTGATGCCATGGAATTGAATC 464


RESULT 15
B52785
LOCUS       B52785                   540 bp    DNA     linear   GSS 20-JUN-1998
DEFINITION  CIT-HSP-2006O4.TR CIT-HSP Homo sapiens genomic clone 2006O4,
            genomic survey sequence.
ACCESSION   B52785
VERSION     B52785.1  GI:2607119
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CIT-HSP-2006O4.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7039830"
                     /db_xref="taxon:9606"
                     /clone="2006O4"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      189 a    103 c     69 g    179 t
ORIGIN
```

```
  Query Match             75.8%;  Score 18.2;  DB 28;  Length 540;
  Best Local Similarity   87.0%;  Pred. No. 9.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GTAGTGATGCCATGTAACTGAAT 23
              | |||||| ||||||| |||||| .
Db         81 GAAGTGATTCCATGTATCTGAAT 103


Search completed: February 16, 2004, 13:41:05
Job time : 97.2277 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:55:00 ; Search time 11.7756 Seconds
(without alignments)
5501.769 Million cell updates/sec

Title: US-10-676-079-5
Perfect score: 24
Sequence: 1 tatgatcctctagtacttctcgac 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	100.0	24	21	AAZ39197	Human heparanase P
	2	24	100.0	24	21	AAZ33292	Human heparanase P
c	3	24	100.0	230	24	ABL57029	Nuclear-targeted l
c	4	24	100.0	230	25	ABV77225	Nucleotide sequenc
c	5	24	100.0	432	24	ABK49828	Plasmid pFastBac1-
c	6	24	100.0	604	22	AAA88918	Nucleotide sequenc
c	7	24	100.0	860	22	AAA88917	Nucleotide sequenc
c	8	24	100.0	956	22	AAA88916	Nucleotide sequenc
c	9	24	100.0	1189	24	AAI64290	Protease D-G catay
c	10	24	100.0	1430	22	AAH74721	Nucleotide sequenc
c	11	24	100.0	2062	22	AAA88914	Nucleotide sequenc
c	12	24	100.0	2062	22	AAA88915	Nucleotide sequenc
c	13	24	100.0	2436	22	AAA88921	Nucleotide sequenc
c	14	24	100.0	2436	22	AAA88924	Nucleotide sequenc
c	15	24	100.0	2692	22	AAA88920	Nucleotide sequenc
c	16	24	100.0	2692	22	AAA88923	Nucleotide sequenc
c	17	24	100.0	2788	22	AAA88919	Nucleotide sequenc
c	18	24	100.0	2788	22	AAA88922	Nucleotide sequenc
c	19	24	100.0	6526	21	AAC55471	Destination vector
c	20	24	100.0	6613	21	AAC55493	Destination vector
c	21	24	100.0	6708	21	AAC55475	Destination vector
c	22	24	100.0	7066	21	AAC55498	Destination vector
c	23	24	100.0	7668	24	ABK49827	Plasmid pFastBac1-
c	24	24	100.0	8435	24	ABK49825	Plasmid pFastBac1-
c	25	24	100.0	8505	24	ABK49826	Plasmid pFastBac1-
c	26	23	95.8	934	22	AAF30277	Mouse CD40 recepto
c	27	23	95.8	6668	21	AAC55495	Destination vector
c	28	18.2	75.8	173810	24	ABN85752	Mouse chromosome 1
	29	17.4	72.5	418	24	ABK95198	PCR template conta
	30	17.4	72.5	418	24	ABK95199	PCR template conta
c	31	17.2	71.7	3755	21	AAA30570	Genomic DNA encodi
c	32	17.2	71.7	3755	24	ABL55698	Candida tropicalis
c	33	17.2	71.7	3755	24	ABK31889	Candida tropicalis
c	34	16.6	69.2	271	25	ABX28810	Human GDP-mannose
c	35	16.6	69.2	765	20	AAV69545	P. neopolydactyla
c	36	16.6	69.2	6393	20	AAV70137	TutC and TutB gene
c	37	16.6	69.2	6393	22	AAF23624	TutB/C coding sequ
	38	16.6	69.2	35000	24	ABT12817	Human RECQL gene-r
c	39	16.6	69.2	335913	22	AAI61371	Soybean 240017 reg
c	40	16.6	69.2	335913	22	AAI61372	Soybean 240017 reg
c	41	16.6	69.2	349980	22	AAH41225	Pyrococcus abyssi
	42	16.2	67.5	144	25	ABZ78549	Tumour suppression
	43	16.2	67.5	144	25	ABZ09096	Human oligonucleot
c	44	16.2	67.5	325	22	AAS32313	Human cDNA encodin
	45	16.2	67.5	407	22	AAI91021	Human polynucleoti

ALIGNMENTS

RESULT 1
AAZ39197
ID AAZ39197 standard; DNA; 24 BP.
XX
AC AAZ39197;
XX
DT 02-MAR-2000 (first entry)
XX
DE Human heparanase PCR antisense primer SEQ ID NO:5.
XX
KW Human; heparanase; hpa; genetic modification; expression; anticancer;
KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW micrometastasis; autoimmune lesion; kidney failure; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9957244-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09256.
XX
PR 01-MAY-1998; 98US-0071618.
PR 02-MAR-1999; 99US-0260038.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (FRIE/) FRIEDMAN M M.
XX
PI Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I, Peleg Y;
PI Shlomi Y;
XX
DR WPI; 2000-062144/05.
XX
PT Engineered cells that express recombinant heparanase, useful
PT therapeutically, e.g. for treating angiogenesis and to screen for
PT specific inhibitors, potential anticancer agents -
XX
PS Example 1; Page 36; 118pp; English.
XX
CC The present invention describes genetically modified cells (A) containing
CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,
CC and express recombinant heparanase (II). Heparanase cleaves heparan
CC sulphate (HS) at specific intrachain sites, resulting in release of
CC heparin-binding growth factors, enzymes and proteins that are sequestered
CC by HS in basement membranes, extracellular matrix or cell surfaces. It
CC may also be implicated in tumour angiogenesis and metastases. (II) is
CC potentially useful in wound healing and for treating angiogenesis,
CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
CC infection and cystic fibrosis. It can also be used to neutralise heparin
CC (an alternative to protamine) and to screen for specific inhibitors
CC (potentially useful for treating cancer and metastases). Antibodies
CC raised against (II) are used for immunodetection and diagnosis of
CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
CC in large quantities, in a form that is homogeneously processed and
CC activated/neutralised by a dedicated protease. The present sequence
CC represents a PCR primer for human heparanase, which is used in an
CC example from the present invention.
XX
SQ Sequence 24 BP; 5 A; 7 C; 3 G; 9 T; 0 other;

Query Match 100.0%; Score 24; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCCTCTAGTACTTCTCGAC 24
||||||||||||||||||||||||
Db 1 TATGATCCTCTAGTACTTCTCGAC 24

RESULT 2
AAZ33292
ID AAZ33292 standard; DNA; 24 BP.
XX
AC AAZ33292;
XX
DT 21-FEB-2000 (first entry)
XX
DE Human heparanase PCR antisense primer SEQ ID NO:5.
XX
KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW autoimmune disease; anticancer; kidney disease; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09255.
XX
PR 01-MAY-1998; 98US-0071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
XX
PT Heparanase-specific molecular probes useful for diagnosis and
PT treatment, e.g. of tumors, and for targeted drug delivery -
XX
PS Example; Page 27; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and
CC anti-heparanase antibodies are used to detect or quantify the expression
CC of heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary,
CC cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC syndrome, sepsis and inflammatory or autoimmune disease), for targeted
CC drug delivery (e.g. of anticancer agents) and as research reagents.
CC The present sequence represents a PCR primer for human heparanase, which
CC is used in an example from the present invention for the construction of
CC a heparanase expression vector.
XX
SQ Sequence 24 BP; 5 A; 7 C; 3 G; 9 T; 0 other;

Query Match 100.0%; Score 24; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCCTCTAGTACTTCTCGAC 24
||||||||||||||||||||||||
Db 1 TATGATCCTCTAGTACTTCTCGAC 24

RESULT 3
ABL57029/c
ID ABL57029 standard; DNA; 230 BP.
XX
AC ABL57029;
XX
DT 23-JUL-2002 (first entry)
XX
DE Nuclear-targeted lacZ baculovirus transfection cassette.
XX
KW Baculovirus; vector; gene therapy; lacZ; beta-galactosidase;
KW enzyme; vasotropic; antiatherosclerotic; thrombolytic; cytostatic;
KW CMV; pFASTBac1; ss.
XX
OS Baculovirus.
XX
FH Key Location/Qualifiers
FT misc_feature 6
FT /*tag= a
FT /note= "transcription start for the polyhedrin
FT promoter"
XX
PN WO200190390-A1.
XX
PD 29-NOV-2001.
XX
PF 29-MAY-2001; 2001WO-GB02383.
XX
PR 26-MAY-2000; 2000GB-0012997.
XX
PA (ARKT-) ARK THERAPEUTICS LTD.
XX
PI Yla-Herttuala S, Airenne KJ;
XX
DR WPI; 2002-401582/43.

XX
PT Delivery of gene product, comprises applying the gene in baculovirus
PT vector to blood-free body compartment, useful e.g. for gene therapy of
PT vascular diseases using periadventitial collar -
XX
PS Example 2; Fig 1; 20pp; English.
XX
CC The present sequence is a portion of baculovirus transfer vector
CC pFASTBac1 (pFB). A cytomegalovirus nuclear-targeted lacZ
CC expression cassette (see ABL57028) was inserted into the vector at
CC an StuI site (nucleotides 121-126 of the present sequence), in
CC opposite orientation to the vector's polyhedrin promoter, to
CC construct a nuclear-targeted beta-galactosidase-encoding baculovirus
CC vector. The vector was used in examples from the invention to
CC demonstrate baculovirus-mediated gene transfer to the rabbit
CC arterial wall, rat brain and rabbit skeletal muscle. The invention
CC relates to the delivery of a gene product by providing the gene in
CC a baculovirus vector and applying the vector to a body compartment
CC which is free (or usually free) of blood. In a device for the
CC periadventitial delivery of a gene product, the gene is provided in
CC a baculovirus vector from which the gene is expressed. The
CC baculovirus vector is useful for gene therapy, especially of
CC vascular diseases such as post-angioplasty restenosis, post-bypass
CC atherosclerosis, stenosis of vascular prosthesis anastomoses and
CC thrombus formation. It is also useful for administration to the
CC brain (e.g. for cancer treatment) or ex vivo injection into
CC saline-perfused organs or vessels for transplantation.
CC Note: The specification was published without claims.
XX
SQ Sequence 230 BP; 64 A; 56 C; 53 G; 57 T; 0 other;

Query Match 100.0%; Score 24; DB 24; Length 230;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCCTCTAGTACTTCTCGAC 24
||||||||||||||||||||||||
Db 225 TATGATCCTCTAGTACTTCTCGAC 202

RESULT 4
ABV77225/c
ID ABV77225 standard; DNA; 230 BP.
XX
AC ABV77225;
XX
DT 28-MAR-2003 (first entry)
XX
DE Nucleotide sequence of a CMV-nt lacZ expression cassette fragment.
XX
KW Baculovirus; vector; spinal cord; central nervous system; CNS;
KW periadventitial gene transfer; enzyme replacement; gene therapy;
KW subarachnoid haemorrhage; gene delivery; cerebral choroid plexus;
KW brain disorder; cancer; brain cancer; lacZ; ss.
XX
OS Synthetic.
XX
PN WO200296469-A2.
XX
PD 05-DEC-2002.
XX
PF 28-MAY-2002; 2002WO-GB02504.
XX
PR 29-MAY-2001; 2001WO-GB02383.
PR 29-NOV-2001; 2001GB-0028620.
XX
PA (ARKT-) ARK THERAPEUTICS LTD.
XX
PI Yla-herttuala S, Airenne KJ, Lehtolainen P;
XX
DR WPI; 2003-129500/12.
XX
PT Use of baculovirus vector containing a gene, for the manufacture of a
PT medicament for treating a condition mediated by spinal cord or central
PT nervous system, by the action of a gene or its product -
XX
PS Disclosure; Fig 1; 14pp; English.
XX
CC The specification describes the use of a baculovirus vector containing
CC a gene, for the manufacture of a medicament for the treatment of a
CC condition that can be mediated via the spinal cord or central nervous
CC system (CNS), by the action of a gene or its product. Baculoviruses are
CC capable of mediating periadventitial gene transfer. The vector is useful
CC for treating a condition that requires enzyme replacement such as
CC subarachnoid haemorrhage. It is useful as an efficient tool for gene
CC delivery to cerebral choroid plexus cells and in gene therapy of several
CC types of brain disorders. It is also useful for treating cancer e.g. in
CC brain. To analyse the gene transfer efficiency of baculovirus and
CC adenovirus vectors (both comprising the LacZ cassette), the viruses were
CC injected into corpus callosum of adult rats. Expression of LacZ was
CC analysed by reverse transcriptase (RT)-polymerase chain reaction (PCR).
CC The present sequence represents a CMV-nt lacZ expression cassette
CC fragment, used in adenovirus or baculovirus vectors.
XX
SQ Sequence 230 BP; 64 A; 56 C; 53 G; 57 T; 0 other;

Query Match 100.0%; Score 24; DB 25; Length 230;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCCTCTAGTACTTCTCGAC 24
||||||||||||||||||||||||
Db 225 TATGATCCTCTAGTACTTCTCGAC 202

RESULT 5
ABK49828/c
ID ABK49828 standard; DNA; 432 BP.
XX
AC ABK49828;
XX
DT 15-JUL-2002 (first entry)
XX
DE Plasmid pFastBac1-HT region encoding MycHis tag.
XX
KW ds; ADAMTS; cytostatic; antidiabetic; antirheumatic; MycHis tag;
KW antiarthritic; antiulcer; vulnerary; neovascularisation; angioma;
KW diabetic omentopathy; chronic rheumatoid arthritis; gene therapy;
KW refractory skin ulcer; gastric ulcer; post-operative healing failure;
KW reprolysin-type ZN-metalloprotease domain; disintegrin-like domain; TSP1;
KW thrombospondin type 1 domain; sexual cycle; tumour; 5P-syndrome deletion;
KW chromosome 5p15.2-15.3; Cri-du-chat syndrome; pFastBac1-HT.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 255..320
FT /*tag= a
FT /product= "MycHis tag"
FT /note= "No start or stop codon"
XX
PN WO200231163-A1.
XX
PD 18-APR-2002.
XX
PF 11-OCT-2001; 2001WO-JP08913.
XX
PR 11-OCT-2000; 2000JP-0311309.
PR 02-APR-2001; 2001JP-0102905.
XX
PA (KAZU-) KAZUSA DNA RES INST FOUND.
PA (MITS-) MITSUBISHI PHARMA CORP.
XX
PI Ohara O, Nagase T, Nomura N, Yano K, Murakami K, Yasuda S;
PI Kanzaki K;
XX

```
DR   WPI; 2002-372277/40.
DR   P-PSDB; AAU80152.
XX
PT   Human brain-originated ADAMTS family polypeptide and encoded gene,
PT   applicable in diagnosis and screening compounds for drug compositions
PT   in treating diseases due to e.g. neovascularisation -
XX
PS   Example 20; Fig 16; 172pp; Japanese.
XX
CC   The invention relates to a polypeptide belonging to the ADAMTS family is
CC   selected from sequences appearing as AAU79496, AAU79497 and AAU79499,
CC   a protein that contains the polypeptide, a protein having not less than
CC   50% homology with the amino acid sequence of the polypeptides or a
CC   polypeptide modified from any of the polypeptides but with some amino
CC   acids deleted, substituted, added or inserted. Also included are the
CC   polynucleotides encoding the polypeptides (or their complementary
CC   strands or variants), a recombinant vector containing any of the
CC   polynucleotides, a transformant which is transformed with the recombinant
CC   vector, producing the polypeptide, protein or peptide by culturing the
CC   transformant, an antibody that can recognize the polypeptide, protein or
CC   peptide and screening compounds to promote or inhibit activity of the
CC   polypeptide or protein, or to promote or inhibit expression of the
CC   polynucleotide by using the polypeptide, protein, peptide, the
CC   polynucleotide, vector, transformant or/and antibody, particularly in
CC   the presence of a test compound for contact before evaluating the
CC   activity by measuring signal changes. The polypeptide and encoded gene
CC   are applicable in diagnosis and screening compounds for drug compositions
CC   in treating diseases due to neovascularisation, diabetic omentopathy,
CC   chronic rheumatoid arthritis, angioma, refractory skin and gastric ulcers
CC   and post-operative healing failure, including gene therapy.
CC   The gene encoding such polypeptide has conserved reprolysin-type ZN-
CC   metalloprotease domain, disintegrin-like domain and TSP1 (thrombospondin
CC   type 1) domain. Its encoded protein is characterised by high expression
CC   in ovaries, changes in expression dose depending on the sexual cycle, a
CC   decrease in tumour cell and location of the gene on the 5P-syndrome
CC   deletion site on chromosome 5p15.2-15.3 (associated with
CC   Cri-du-chat syndrome). The present sequence is a region of plasmid
CC   pFastBac1-HT used to make the expression Plasmid pFastBac1-HT-PJ01256
CC   used to express DNA encoding an ADAMTS protein.
XX
SQ   Sequence 432 BP; 138 A; 105 C; 82 G; 107 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 24;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 0.17;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db        350 TATGATCCTCTAGTACTTCTCGAC 327


RESULT 6
AAA88918/c
ID   AAA88918 standard; DNA; 604 BP.
XX
AC   AAA88918;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Nucleotide sequence of ricin toxin B chain RTB3 in pFASTBAC1.
XX
KW   Ricin toxin B; RTB3; crystal protein; insecticide; pesticide;
KW   toxin; transgenic plant; insect resistance; crop protection;
KW   lectin; pFASTBAC1; ds.
XX
OS   Chimeric - Ricinus communis.
OS   Chimeric - Baculovirus.
XX
PN   WO200066755-A2.
XX
PD   09-NOV-2000.
XX
PF   27-APR-2000; 2000WO-GB01633.
XX
PR   28-APR-1999;  99GB-0009796.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
XX
DR   WPI; 2001-007228/01.
XX
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests -
XX
PS   Claim 10; Fig 3e; 81pp; English.
XX
CC   The present sequence is that of ricin toxin B chain RTB3 DNA in
CC   baculovirus transfer vector pFASTBAC1, in which the DNA was cloned
CC   under the control of the polyhedrin promoter. The invention
CC   provides nucleic acids (see AAA88919-24) encoding pesticidal fusion
CC   proteins comprising a toxin domain and a heterologous binding domain
CC   capable of binding non-specifically to a cell membrane without
CC   disrupting that membrane. The toxin domain is preferably obtained
CC   from Bacillus thuringiensis crystal proteins CryIA(b) or CryIA(c),
CC   and the binding domain is preferably derived from a lectin, especially
CC   ricin toxin B chain. The use of such fusions may help to inhibit
CC   the acquisition of resistance in a pest population treated with the
CC   protein. Vectors (e.g. baculovirus vectors or vectors suitable for
CC   use in a plant), host cells, and transgenic plants (especially rice
CC   or maize) are also provided. Expression of the fusion protein in a
CC   plant is useful for influencing or affecting the toxicity of a
CC   plant to a pest, allowing control of e.g. Lepidoptera, Coleoptera,
CC   Culicidae, Simuliidae, Hymenoptera, Homoptera, Diptera and
CC   Orthoptera pests.
XX
SQ   Sequence 604 BP; 180 A; 121 C; 134 G; 169 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 22;  Length 604;
  Best Local Similarity   100.0%;  Pred. No. 0.17;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db        599 TATGATCCTCTAGTACTTCTCGAC 576


RESULT 7
AAA88917/c
ID   AAA88917 standard; DNA; 860 BP.
XX
AC   AAA88917;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Nucleotide sequence of ricin toxin B chain RTB2 in pFASTBAC1.
XX
KW   Ricin toxin B; RTB2; crystal protein; insecticide; pesticide;
KW   toxin; transgenic plant; insect resistance; crop protection;
KW   lectin; pFASTBAC1; ds.
XX
OS   Chimeric - Ricinus communis.
OS   Chimeric - Baculovirus.
XX
PN   WO200066755-A2.
XX
PD   09-NOV-2000.
XX
PF   27-APR-2000; 2000WO-GB01633.
XX
PR   28-APR-1999;  99GB-0009796.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
```

```
XX
DR   WPI; 2001-007228/01.
XX
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests -
XX
PS   Claim 10; Fig 3d; 81pp; English.
XX
CC   The present sequence is that of ricin toxin B chain RTB2 DNA in
CC   baculovirus transfer vector pFASTBAC1, in which the DNA was cloned
CC   under the control of the polyhedrin promoter. The invention
CC   provides nucleic acids (see AAA88919-24) encoding pesticidal fusion
CC   proteins comprising a toxin domain and a heterologous binding domain
CC   capable of binding non-specifically to a cell membrane without
CC   disrupting that membrane. The toxin domain is preferably obtained
CC   from Bacillus thuringiensis crystal proteins CryIA(b) or CryIA(c),
CC   and the binding domain is preferably derived from a lectin, especially
CC   ricin toxin B chain. The use of such fusions may help to inhibit
CC   the acquisition of resistance in a pest population treated with the
CC   protein. Vectors (e.g. baculovirus vectors or vectors suitable for
CC   use in a plant), host cells, and transgenic plants (especially rice
CC   or maize) are also provided. Expression of the fusion protein in a
CC   plant is useful for influencing or affecting the toxicity of a
CC   plant to a pest, allowing control of e.g. Lepidoptera, Coleoptera,
CC   Culicidae, Simuliidae, Hymenoptera, Homoptera, Diptera and
CC   Orthoptera pests.
XX
SQ   Sequence 860 BP; 256 A; 164 C; 199 G; 241 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 22;  Length 860;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db        855 TATGATCCTCTAGTACTTCTCGAC 832


RESULT 8
AAA88916/c
ID   AAA88916 standard; DNA; 956 BP.
XX
AC   AAA88916;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Nucleotide sequence of ricin toxin B chain RTB1 in pFASTBAC1.
XX
KW   Ricin toxin B; RTB1; crystal protein; insecticide; pesticide;
KW   toxin; transgenic plant; insect resistance; crop protection;
KW   lectin; pFASTBAC1; ds.
XX
OS   Chimeric - Ricinus communis.
OS   Chimeric - Baculovirus.
XX
PN   WO200066755-A2.
XX
PD   09-NOV-2000.
XX
PF   27-APR-2000; 2000WO-GB01633.
XX
PR   28-APR-1999;   99GB-0009796.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
XX
DR   WPI; 2001-007228/01.
XX
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests -
XX
PS   Claim 10; Fig 3c; 81pp; English.
XX
CC   The present sequence is that of ricin toxin B chain RTB1 DNA in
CC   baculovirus transfer vector pFASTBAC1, in which the DNA was cloned
CC   under the control of the polyhedrin promoter. The invention
CC   provides nucleic acids (see AAA88919-24) encoding pesticidal fusion
CC   proteins comprising a toxin domain and a heterologous binding domain
CC   capable of binding non-specifically to a cell membrane without
CC   disrupting that membrane. The toxin domain is preferably obtained
CC   from Bacillus thuringiensis crystal proteins CryIA(b) or CryIA(c),
CC   and the binding domain is preferably derived from a lectin, especially
CC   ricin toxin B chain. The use of such fusions may help to inhibit
CC   the acquisition of resistance in a pest population treated with the
CC   protein. Vectors (e.g. baculovirus vectors or vectors suitable for
CC   use in a plant), host cells, and transgenic plants (especially rice
CC   or maize) are also provided. Expression of the fusion protein in a
CC   plant is useful for influencing or affecting the toxicity of a
CC   plant to a pest, allowing control of e.g. Lepidoptera, Coleoptera,
CC   Culicidae, Simuliidae, Hymenoptera, Homoptera, Diptera and
CC   Orthoptera pests.
XX
SQ   Sequence 956 BP; 286 A; 188 C; 214 G; 268 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 22;  Length 956;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db        951 TATGATCCTCTAGTACTTCTCGAC 928


RESULT 9
AAI64290/c
ID   AAI64290 standard; DNA; 1189 BP.
XX
AC   AAI64290;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Protease D-G cataytic domain fusion gene construct encoding sequence.
XX
KW   Serine protease; D-G; human; zymogen; enzyme; cytostatic;
KW   antiinflammatory; dermatological; anticoagulation; cancer;
KW   skin disorder; neuropathic pain; inflammatory disorder;
KW   coagulation diathesis; thrombosis; laundry detergent; skin care;
KW   gene therapy; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             13..891
FT                   /*tag=  a
FT                   /product= "protease D-G"
FT                   /transl_except= (pos: 640..642, aa: Gln)
FT   sig_peptide     13..93
FT                   /*tag=  b
FT                   /note= "Prolactin signal sequence"
FT   mat_peptide     94..888
FT                   /*tag=  c
FT                   /note= "Mature protease D-G"
XX
PN   WO200202011-A1.
XX
PD   10-JAN-2002.
XX
PF   08-JUN-2001; 2001WO-US18568.
XX
PR   30-JUN-2000; 2000US-0607745.
XX
PA   (ORTH ) ORTHO-MCNEIL PHARM INC.
XX
```

PI Darrow AL, Qi J, Andrade-Gordon P;
XX
DR WPI; 2002-106601/14.
DR P-PSDB; AAG78578.
XX
PT Nucleic acid encoding a serine protease called D-G protein which is
PT useful for identifying modulators that are useful for treating a
PT condition which is mediated by protease D-G, e.g. cancer, skin
PT disorders, or neuropathic pain -
XX
PS Claim 4; Fig 4B; 81pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid that
CC encodes a serine protease called D-G protein. The activity of the protein
CC of the invention may be described as cytostatic, antiinflammatory,
CC dermatological and anticoagulation. The serine protease of the invention
CC is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
CC which play an important role in processes such as digestion and
CC regulatory amplification cascades through the proteolytic activation of
CC inactive zymogen precursors. Protease D-G modulating compounds are useful
CC for treating a condition which is mediated by protease D-G, e.g. cancer,
CC skin disorders, neuropathic pain, inflammatory disorders, or coagulation
CC diathesis/thrombosis. The polynucleotide encoding the protease is useful
CC for identifying, detecting or isolating mutant forms of DNA molecules
CC encoding the protease. The protease is useful for identifying modulators
CC of the functional protease. The D-G protein can be used for formulation
CC of compositions for laundry detergents and skin care products. Protease
CC D-G gene therapy may be used to introduce protease D-G into the cells of
CC target organisms. As the D-G protein is derived from a human, it is less
CC likely to produce an allergic reaction in sensitive individuals when used
CC in formulations for laundry detergents and skin care products. The
CC current sequence represents the protease D-G cataytic domain in the
CC zymogen activation construct encoding sequence.
XX
SQ Sequence 1189 BP; 297 A; 305 C; 307 G; 280 T; 0 other;

Query Match 100.0%; Score 24; DB 24; Length 1189;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCCTCTAGTACTTCTCGAC 24
||||||||||||||||||||||||
Db 1087 TATGATCCTCTAGTACTTCTCGAC 1064

RESULT 10
AAH74721/c
ID AAH74721 standard; DNA; 1430 BP.
XX
AC AAH74721;
XX
DT 29-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a leech active polypeptide fragment.
XX
KW Active polypeptide; ectoparasitic leech; Rhynchobdellida; C3bBb complex;
KW complement activation; complement factor D; haemodialysis; catheter;
KW cardio-pulmonary bypass; extra-arterial stent; transplant rejection;
KW autoimmune disease; lupus arthritis; rheumatoid arthritis; sepsis;
KW glomerulonephritis; nephritis; nephropathy; reperfusion; anaphylaxis;
KW asthma; skin reaction; infection; sickle cell anemia; haemolytic anemia;
KW ds.
XX
OS Placobdella papillifera.
XX
PN WO200147963-A2.
XX
PD 05-JUL-2001.
XX
PF 21-DEC-2000; 2000WO-GB04971.
XX
PR 24-DEC-1999; 99GB-0030659.
XX

PA (BIOD-) BIO-DISCOVERY LTD.
XX
PI Finney S, Seale L, Wallis RB;
XX
DR WPI; 2001-522011/57.
XX
PT Novel polypeptides from the leech Placobdella papillifera which inhibit
PT alternate pathway of complement activation, useful for treating
PT rheumatoid arthritis, sepsis, asthma involving alternate complement
PT pathway -
XX
PS Claim 18; Page 53-54; 80pp; English.
XX
CC The present sequence encodes an active polypeptide fragment. The active
CC polypeptide has a molecular weight of 7000-17,000 Da (as measured by
CC mass spectrometry), and is derived from ectoparasitic leeches, of order
CC Rhynchobdellida, of genus Placobdella and especially of species
CC P. papillifera. The polypeptide inhibits the alternate route of
CC complement activation but has substantially no effect on complement
CC activation by the classical route. The polypeptide is an inhibitor of
CC complement factor D and/or C3bBb complex. The active polypeptide is
CC useful for manufacturing a medicament and inhibits one or more steps
CC in the alternate pathway of complement activation. It is useful for
CC treating or preventing conditions, such as haemodialysis and
CC cardio-pulmonary bypass, the presence of in-dwelling catheters and
CC extra-arterial stents, rejection of transplanted organs or tissues,
CC autoimmune diseases including lupus arthritis, rheumatoid arthritis,
CC glomerulonephritis, nephritis, nephropathy, sepsis, injury caused to
CC tissues by reperfusion after an ischaemic period and other conditions
CC associated with activation of complement, including anaphylaxis,
CC asthma, skin reactions, infections, sickle cell anemia and haemolytic
CC anemia involving activation of alternate complement pathway in a
CC patient.
XX
SQ Sequence 1430 BP; 419 A; 296 C; 296 G; 419 T; 0 other;

Query Match 100.0%; Score 24; DB 22; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCCTCTAGTACTTCTCGAC 24
||||||||||||||||||||||||
Db 1126 TATGATCCTCTAGTACTTCTCGAC 1103

RESULT 11
AAA88914/c
ID AAA88914 standard; DNA; 2062 BP.
XX
AC AAA88914;
XX
DT 05-MAR-2001 (first entry)
XX
DE Nucleotide sequence of CryIA(b) in pFASTBAC1.
XX
KW CryIA(b); crystal protein; insecticide; pesticide; toxin;
KW transgenic plant; insect resistance; crop protection; pFASTBAC1;
KW ds.
XX
OS Chimeric - Bacillus thuringiensis.
OS Chimeric - Baculovirus.
XX
FH Key Location/Qualifiers
FT CDS 97..1944
FT /*tag= a
XX
PN WO200066755-A2.
XX
PD 09-NOV-2000.
XX
PF 27-APR-2000; 2000WO-GB01633.
XX
PR 28-APR-1999; 99GB-0009796.

```
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
XX
DR   WPI; 2001-007228/01.
XX
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests  -
XX
PS   Claim 9; Fig 3a; 81pp; English.
XX
CC   The present sequence is that of Bacillus thuringiensis crystal
CC   protein CryIA(b) DNA in baculovirus transfer vector pFASTBAC1, in
CC   which the gene was cloned under the control of the polyhedrin
CC   promoter.  The invention provides nucleic acids (see AAA88919-24)
CC   encoding pesticidal fusion proteins comprising a toxin domain and a
CC   heterologous binding domain capable of binding non-specifically to
CC   a cell membrane without disrupting that membrane.  The toxin domain
CC   is preferably obtained from CryIA(b) or CryIA(c), and the binding
CC   domain is preferably derived from a lectin, such as ricin toxin B
CC   chain.  The use of such fusions may help to inhibit the acquisition
CC   of resistance in a pest population treated with the protein.
CC   Vectors (e.g. baculovirus vectors or vectors suitable for use in a
CC   plant), host cells, and transgenic plants (especially rice or
CC   maize) are also provided.  Expression of the fusion protein in a
CC   plant is useful for influencing or affecting the toxicity of a
CC   plant to a pest, allowing control of e.g. Lepidoptera, Coleoptera,
CC   Culicidae, Simuliidae, Hymenoptera, Homoptera, Diptera and
CC   Orthoptera pests.
XX
SQ   Sequence 2062 BP; 538 A; 552 C; 440 G; 532 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 22;  Length 2062;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2057 TATGATCCTCTAGTACTTCTCGAC 2034


RESULT 12
AAA88915/c
ID   AAA88915 standard; DNA; 2062 BP.
XX
AC   AAA88915;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Nucleotide sequence of CryIA(c) in pFASTBAC1.
XX
KW   CryIA(c); crystal protein; insecticide; pesticide; toxin;
KW   transgenic plant; insect resistance; crop protection; pFASTBAC1;
KW   ds.
XX
OS   Chimeric - Bacillus thuringiensis.
OS   Chimeric - Baculovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             97..1944
FT                   /*tag=  a
XX
PN   WO200066755-A2.
XX
PD   09-NOV-2000.
XX
PF   27-APR-2000; 2000WO-GB01633.
XX
PR   28-APR-1999;   99GB-0009796.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
XX
DR   WPI; 2001-007228/01.
XX
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests  -
XX
PS   Claim 9; Fig 3b; 81pp; English.
XX
CC   The present sequence is that of Bacillus thuringiensis crystal
CC   protein CryIA(c) DNA in baculovirus transfer vector pFASTBAC1, in
CC   which the gene was cloned under the control of the polyhedrin
CC   promoter.  The invention provides nucleic acids (see AAA88919-24)
CC   encoding pesticidal fusion proteins comprising a toxin domain and a
CC   heterologous binding domain capable of binding non-specifically to
CC   a cell membrane without disrupting that membrane.  The toxin domain
CC   is preferably obtained from CryIA(b) or CryIA(c), and the binding
CC   domain is preferably derived from a lectin, such as ricin toxin B
CC   chain.  The use of such fusions may help to inhibit the acquisition
CC   of resistance in a pest population treated with the protein.
CC   Vectors (e.g. baculovirus vectors or vectors suitable for use in a
CC   plant), host cells, and transgenic plants (especially rice or
CC   maize) are also provided.  Expression of the fusion protein in a
CC   plant is useful for influencing or affecting the toxicity of a
CC   plant to a pest, allowing control of e.g. Lepidoptera, Coleoptera,
CC   Culicidae, Simuliidae, Hymenoptera, Homoptera, Diptera and
CC   Orthoptera pests.
XX
SQ   Sequence 2062 BP; 544 A; 532 C; 446 G; 540 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 22;  Length 2062;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2057 TATGATCCTCTAGTACTTCTCGAC 2034


RESULT 13
AAA88921/c
ID   AAA88921 standard; DNA; 2436 BP.
XX
AC   AAA88921;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Nucleotide sequence of CryIA(b)-RTB3 fusion in pFASTBAC1.
XX
KW   CryIA(b); crystal protein; ricin toxin B; RTB3; lectin;
KW   insecticide; pesticide; toxin; transgenic plant; insect resistance;
KW   crop protection; pFASTBAC1; ds.
XX
OS   Chimeric - Bacillus thuringiensis.
OS   Chimeric - Ricinus communis.
OS   Chimeric - Baculovirus.
XX
PN   WO200066755-A2.
XX
PD   09-NOV-2000.
XX
PF   27-APR-2000; 2000WO-GB01633.
XX
PR   28-APR-1999;   99GB-0009796.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
XX
DR   WPI; 2001-007228/01.
XX
```

```
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests  -
XX
PS   Claim 11; Fig 3h; 81pp; English.
XX
CC   This nucleotide sequence is that of a fusion between DNA encoding
CC   crystal protein CryIA(b) (see AAA88914) of Bacillus thuringiensis and
CC   DNA encoding ricin toxin B RTB3 (see AAA88918) in baculovirus transfer
CC   vector pFASTBAC1, in which the fusion was cloned under the control
CC   of the polyhedrin promoter.  This is an example of claimed nucleic
CC   acids encoding pesticidal fusion proteins between a toxin domain
CC   and a heterologous binding domain capable of binding non-specifically
CC   to a cell membrane without disrupting that membrane.  The use of such
CC   fusions may help to inhibit the acquisition of resistance in a pest
CC   population treated with the protein.  Vectors (e.g. baculovirus
CC   vectors or vectors suitable for use in a plant), host cells, and
CC   transgenic plants (especially rice or maize) are also provided.
CC   Expression of the fusion protein in a plant is useful for influencing
CC   or affecting the toxicity of a plant to a pest, allowing control of
CC   e.g. Lepidoptera, Coleoptera, Culicidae, Simuliidae, Hymenoptera,
CC   Homoptera, Diptera and Orthoptera pests.
XX
SQ   Sequence 2436 BP; 654 A; 617 C; 521 G; 644 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 22;  Length 2436;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2431 TATGATCCTCTAGTACTTCTCGAC 2408


RESULT 14
AAA88924/c
ID   AAA88924 standard; DNA; 2436 BP.
XX
AC   AAA88924;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Nucleotide sequence of CryIA(c)-RTB3 fusion in pFASTBAC1.
XX
KW   CryIA(c); crystal protein; ricin toxin B; RTB3; lectin;
KW   insecticide; pesticide; toxin; transgenic plant; insect resistance;
KW   crop protection; pFASTBAC1; ds.
XX
OS   Chimeric - Bacillus thuringiensis.
OS   Chimeric - Ricinus communis.
OS   Chimeric - Baculovirus.
XX
PN   WO200066755-A2.
XX
PD   09-NOV-2000.
XX
PF   27-APR-2000; 2000WO-GB01633.
XX
PR   28-APR-1999;   99GB-0009796.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
XX
DR   WPI; 2001-007228/01.
XX
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests  -
XX
PS   Claim 11; Fig 3k; 81pp; English.
XX
CC   This nucleotide sequence is that of a fusion between DNA encoding
CC   crystal protein CryIA(c) (see AAA88915) of Bacillus thuringiensis and
CC   DNA encoding ricin toxin B RTB3 (see AAA88918) in baculovirus transfer
CC   vector pFASTBAC1, in which the fusion was cloned under the control
CC   of the polyhedrin promoter.  This is an example of claimed nucleic
CC   acids encoding pesticidal fusion proteins between a toxin domain
CC   and a heterologous binding domain capable of binding non-specifically
CC   to a cell membrane without disrupting that membrane.  The use of such
CC   fusions may help to inhibit the acquisition of resistance in a pest
CC   population treated with the protein.  Vectors (e.g. baculovirus
CC   vectors or vectors suitable for use in a plant), host cells, and
CC   transgenic plants (especially rice or maize) are also provided.
CC   Expression of the fusion protein in a plant is useful for influencing
CC   or affecting the toxicity of a plant to a pest, allowing control of
CC   e.g. Lepidoptera, Coleoptera, Culicidae, Simuliidae, Hymenoptera,
CC   Homoptera, Diptera and Orthoptera pests.
XX
SQ   Sequence 2436 BP; 660 A; 597 C; 527 G; 652 T; 0 other;

  Query Match             100.0%;  Score 24;  DB 22;  Length 2436;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2431 TATGATCCTCTAGTACTTCTCGAC 2408


RESULT 15
AAA88920/c
ID   AAA88920 standard; DNA; 2692 BP.
XX
AC   AAA88920;
XX
DT   05-MAR-2001  (first entry)
XX
DE   Nucleotide sequence of CryIA(b)-RTB2 fusion in pFASTBAC1.
XX
KW   CryIA(b); crystal protein; ricin toxin B; RTB2; lectin;
KW   insecticide; pesticide; toxin; transgenic plant; insect resistance;
KW   crop protection; pFASTBAC1; ds.
XX
OS   Chimeric - Bacillus thuringiensis.
OS   Chimeric - Ricinus communis.
OS   Chimeric - Baculovirus.
XX
PN   WO200066755-A2.
XX
PD   09-NOV-2000.
XX
PF   27-APR-2000; 2000WO-GB01633.
XX
PR   28-APR-1999;   99GB-0009796.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Christou P,  Mehlo L;
XX
DR   WPI; 2001-007228/01.
XX
PT   Novel nucleic acid molecule encoding a pesticidal fusion polypeptide
PT   comprising a toxin and a binding domain for producing transgenic plants
PT   resistant to pests  -
XX
PS   Claim 11; Fig 3g; 81pp; English.
XX
CC   This nucleotide sequence is that of a fusion between DNA encoding
CC   crystal protein CryIA(b) (see AAA88914) of Bacillus thuringiensis and
CC   DNA encoding ricin toxin B RTB2 (see AAA88917) in baculovirus transfer
CC   vector pFASTBAC1, in which the fusion was cloned under the control
CC   of the polyhedrin promoter.  This is an example of claimed nucleic
CC   acids encoding pesticidal fusion proteins between a toxin domain
CC   and a heterologous binding domain capable of binding non-specifically
CC   to a cell membrane without disrupting that membrane.  The use of such
```

CC fusions may help to inhibit the acquisition of resistance in a pest
CC population treated with the protein. Vectors (e.g. baculovirus
CC vectors or vectors suitable for use in a plant), host cells, and
CC transgenic plants (especially rice or maize) are also provided.
CC Expression of the fusion protein in a plant is useful for influencing
CC or affecting the toxicity of a plant to a pest, allowing control of
CC e.g. Lepidoptera, Coleoptera, Culicidae, Simuliidae, Hymenoptera,
CC Homoptera, Diptera and Orthoptera pests.
XX
SQ Sequence 2692 BP; 730 A; 660 C; 586 G; 716 T; 0 other;

Query Match 100.0%; Score 24; DB 22; Length 2692;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCCTCTAGTACTTCTCGAC 24
||||||||||||||||||||||||
Db 2687 TATGATCCTCTAGTACTTCTCGAC 2664

Search completed: February 16, 2004, 09:18:11
Job time : 12.7756 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 08:49:50 ; Search time 93.2277 Seconds
(without alignments)
6256.802 Million cell updates/sec

Title: US-10-676-079-5
Perfect score: 24
Sequence: 1 tatgatcctctagtacttctcgac 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		19.2	80.0	820	13	BU371267	BU371267 603596334
2		18.8	78.3	293	12	BI024729	BI024729 RC5-MT025
c 3		18.8	78.3	389	28	AQ105178	AQ105178 HS_3000_A
c 4		18.4	76.7	749	28	AZ805647	AZ805647 2M0067G12
5		18.2	75.8	491	28	AQ502165	AQ502165 V8B1 mTn-
c 6		18.2	75.8	566	9	AA660276	AA660276 00145 MtR
7		18.2	75.8	600	29	BZ309602	BZ309602 ic09c08.b
c 8		18.2	75.8	600	29	BZ324410	BZ324410 ic09c08.g
c 9		18.2	75.8	608	28	AQ305217	AQ305217 HS_2019_A
10		18.2	75.8	609	28	BH022001	BH022001 GH_MBb000
11		18.2	75.8	737	14	CA781659	CA781659 031G11AF
12		18.2	75.8	741	29	BZ506781	BZ506781 BONFA77TF
13		18.2	75.8	799	29	BZ998034	BZ998034 PUGFY17TD
14		18.2	75.8	815	28	AQ862478	AQ862478 nbeb0019A
15		18.2	75.8	1661	29	BZ576392	BZ576392 msh2_4912
16		17.8	74.2	484	9	AW573112	AW573112 hf32c10.x
c 17		17.6	73.3	195	9	AW037530	AW037530 EST278857
c 18		17.6	73.3	327	14	CB891487	CB891487 EST648456
c 19		17.6	73.3	493	9	AI083079	AI083079 MS64t3 Me
20		17.6	73.3	504	9	AI397571	AI397571 NCSC5D12T
c 21		17.6	73.3	533	28	AZ750896	AZ750896 RPCI-24-1
c 22		17.6	73.3	572	28	BH877505	BH877505 hr38e09.b
c 23		17.6	73.3	609	10	BF643933	BF643933 NF092B11E
c 24		17.6	73.3	621	10	BE460948	BE460948 EST412367
c 25		17.6	73.3	640	10	BG449733	BG449733 NF007C08I
c 26		17.6	73.3	645	9	AW586847	AW586847 EST318470
c 27		17.6	73.3	655	10	BF646400	BF646400 NF071F10E
c 28		17.6	73.3	659	13	BQ147235	BQ147235 NF038C01F
c 29		17.6	73.3	695	9	AV403040	AV403040 AV403040
c 30		17.6	73.3	695	12	BP126081	BP126081 BP126081
c 31		17.6	73.3	704	9	AV404024	AV404024 AV404024
c 32		17.6	73.3	740	10	BG581848	BG581848 EST483584
33		17.6	73.3	792	14	CB984130	CB984130 AGENCOURT
34		17.6	73.3	1213	29	CC225588	CC225588 CH261-57B
35		17.4	72.5	415	28	AQ556970	AQ556970 HS_5308_B
36		17.4	72.5	521	29	BZ521578	BZ521578 BOKAJ58TR
c 37		17.4	72.5	728	28	BH995266	BH995266 oeh17c12.
38		17.2	71.7	125	12	BG945612	BG945612 PM2-KN003
39		17.2	71.7	164	10	BF859022	BF859022 MR0-FT018
40		17.2	71.7	188	9	AW064152	AW064152 SP0586 KR
41		17.2	71.7	201	13	BU890760	BU890760 P041C09 P
42		17.2	71.7	235	10	BF858704	BF858704 MR0-FT019
43		17.2	71.7	252	10	BF954669	BF954669 MR0-NN116
44		17.2	71.7	297	10	BF753157	BF753157 MR0-BN038
45		17.2	71.7	378	9	AW181442	AW181442 pa02g07.y

ALIGNMENTS

RESULT 1
BU371267
LOCUS BU371267 820 bp mRNA linear EST 28-NOV-2002
DEFINITION 603596334F1 CSEQCHN73 Gallus gallus cDNA clone ChEST564c9 5', mRNA sequence.
ACCESSION BU371267
VERSION BU371267.1 GI:25879268
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 820)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409

```
          Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST564c9"
                     /sex="Female"
                     /tissue_type="not cerebrum or cerebellum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN73"
                     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with EcoRI
                     , size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
BASE COUNT      204 a    246 c    171 g    199 t
ORIGIN

  Query Match             80.0%;  Score 19.2;  DB 13;  Length 820;
  Best Local Similarity   87.5%;  Pred. No. 5.1e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              |||| | ||||| ||||||||||
Db        381 TATGCTTCTCTACTACTTCTCGAC 404


RESULT 2
BI024729
LOCUS       BI024729                293 bp    mRNA    linear   EST 14-JUN-2001
DEFINITION  RC5-MT0259-300101-012-G04 MT0259 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI024729
VERSION     BI024729.1  GI:14431359
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC5&t2=RC5-MT0259-
            300101-012-G04&t3=2001-01-30&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 48
            High quality sequence stop: 293.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0259"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      70 a     76 c     73 g     74 t
ORIGIN

  Query Match             78.3%;  Score 18.8;  DB 12;  Length 293;
  Best Local Similarity   90.9%;  Pred. No. 5.1e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATGATCCTCTAGTACTTCTCGA 23
              | | ||||||||||||||||||
Db          6 ACGTTCCTCTAGTACTTCTCGA 27


RESULT 3
AQ105178/c
LOCUS       AQ105178                389 bp    DNA     linear   GSS 28-AUG-1998
DEFINITION  HS_3000_A1_H12_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3000 Col=23 Row=O, genomic survey
            sequence.
ACCESSION   AQ105178
VERSION     AQ105178.1  GI:3480534
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 389)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3000  row: O  column: 23
            Class: BAC ends
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1. .389
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3000 Col=23 Row=O"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
```

```
BASE COUNT      122 a     67 c     81 g    108 t     11 others
ORIGIN

  Query Match             78.3%;  Score 18.8;  DB 28;  Length 389;
  Best Local Similarity   90.9%;  Pred. No. 5.7e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATGATCCTCTAGTACTTCTCGA 23
              |||||| |||||||||||||  |
Db        194 ATGATCTTCTAGTACTTCTCTA 173


RESULT 4
AZ805647/c
LOCUS       AZ805647                749 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0067G12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0067G12 F, genomic survey sequence.
ACCESSION   AZ805647
VERSION     AZ805647.1  GI:12966458
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0067  row: G  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 749.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0067G12"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      211 a    203 c    144 g    191 t
ORIGIN

  Query Match             76.7%;  Score 18.4;  DB 28;  Length 749;
  Best Local Similarity   95.0%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCT 20
              | ||||||||||||||||||
Db        529 TCTGATCCTCTAGTACTTCT 510


RESULT 5
AQ502165
LOCUS       AQ502165                491 bp    DNA     linear   GSS 29-APR-1999
DEFINITION  V8B1 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
            genomic 5', genomic survey sequence.
ACCESSION   AQ502165
VERSION     AQ502165.1  GI:4707815
KEYWORDS    GSS.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
            desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatis,D., Jansen,R.,
            Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
            Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
  TITLE     Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
            Gene Disruption
  JOURNAL   Unpublished
COMMENT     Contact: Kumar A
            Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
            Yale University
            P.O. Box 208103, New Haven, CT 06520-8103, USA
            Tel: 203 432 9949
            Fax: 203 432 6161
            Email: anuj.kumar@yale.edu
            te of mTn-3xHA/lacZ insertion.
            Seq primer: GGCCTTCTTTCTTTGGAAGTAC
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /lab_host="E. coli"
                     /clone_lib="mTn-3xHA/lacZ Insertion Library"
                     /note="Vector: pHSS6-Sal; A yeast genomic DNA library
                     (lacking mitochondrial DNA) was prepared in pHSS6-Sal;
                     genomic DNA was size-fractionated (DNA of roughly 2-3 kb
                     in length) prior to cloning.  This library was
                     subsequently mutagenized with a mTn-3xHA/lacZ
                     minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT      125 a    127 c    119 g    116 t      4 others
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 28;  Length 491;
  Best Local Similarity   87.0%;  Pred. No. 1.1e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 ATGATCCTCTAGTACTTCTCGAC 24
              |||||||||| |||||| || ||
Db        107 ATGATCCTCTTGTACTTGTCTAC 129


RESULT 6
AA660276/c
LOCUS       AA660276                566 bp    mRNA    linear   EST 08-MAR-2000
DEFINITION  00145 MtRHE Medicago truncatula cDNA 5' similar to retinoblastoma
            binding protein p46, mRNA sequence.
```

```
ACCESSION   AA660276
VERSION     AA660276.1  GI:2604320
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Covitz,P.A., Smith,L.S. and Long,S.R.
  TITLE     Expressed sequence tags from a root-hair-enriched medicago
            truncatula cDNA library
  JOURNAL   Plant Physiol. 117 (4), 1325-1332 (1998)
COMMENT     Contact: Long SR
            Department of Biological Sciences and Howard Hughes Medical
            Institute
            Stanford University
            Gilbert Biology, Stanford, CA 94305-5020, USA
            Tel: 650 723 3232
            Fax: 650 725 8309
            Email: fa.srl@forsythe.stanford.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /tissue_type="Root hairs & tips"
                     /dev_stage="2-3 day old seedlings"
                     /clone_lib="MtRHE"
                     /note="Organ: Root; Vector: pBK-CMV; Site_1: EcoRI;
                     Site_2: XhoI; cDNA was synthesized from a pooled mRNA prep
                     from elongating root  hairs (30% w/w) and 2-3cm root tips
                     (70% w/w).  XhoI-oligo-dT  linker-primer and EcoRI
                     adaptors were used. cDNAs was cloned  unidirectionally
                     into lambda ZAP Express (Stratagene), amplified,  and
                     mass-excised into pBK-CMV vector plasmids.   More
                     information is available at http://bio-SRL8.stanford.edu."
BASE COUNT      178 a    121 c    135 g    127 t      5 others
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 9;  Length 566;
  Best Local Similarity   83.3%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||  ||| ||||||||  |||
Db        496 TATGATAGTCTTGTACTTCTNGAC 473


RESULT 7
BZ309602
LOCUS       BZ309602                600 bp    DNA     linear   GSS 06-NOV-2002
DEFINITION  ic09c08.b1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
            genomic clone ic09c08 5', genomic survey sequence.
ACCESSION   BZ309602
VERSION     BZ309602.1  GI:24670634
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: ic09  row: c  column: 08
            Seq primer: -21M13UnivFwd
            Class: shotgun
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ic09c08"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"
                     /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                     The vector was digested with XbaI and one nucleotide was
                     added by  fill in in the recessive 3' end. The genomic DNA
                     was nebulized, end  repaired, adaptor ligated and size
                     fractionated using sephadex. The  resulting fragments were
                     between 0.8 and 3 kb and were cloned into  the vector
                     (.x/y reads in M13mp19, .b/g reads in pUC19). The same
                     ligation was transformed in either JM107 or DH5a. "
BASE COUNT      175 a    102 c    134 g    188 t      1 others
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 29;  Length 600;
  Best Local Similarity   87.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGA 23
              ||  |||||||||||||| |||
Db        414 TAATATCCTCTAGTACTTCACGA 436


RESULT 8
BZ324410/c
LOCUS       BZ324410                600 bp    DNA     linear   GSS 06-NOV-2002
DEFINITION  ic09c08.g1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
            genomic clone ic09c08 5', genomic survey sequence.
ACCESSION   BZ324410
VERSION     BZ324410.1  GI:24703958
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: ic09  row: c  column: 08
            Seq primer: -21M13UnivRev
            Class: shotgun
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
```

```
                     /db_xref="taxon:4577"
                     /clone="ic09c08"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"
                     /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                     The vector was digested with XbaI and one nucleotide was
                     added by  fill in in the recessive 3' end. The genomic DNA
                     was nebulized, end  repaired, adaptor ligated and size
                     fractionated using sephadex. The  resulting fragments were
                     between 0.8 and 3 kb and were cloned into  the vector
                     (.x/y reads in M13mp19, .b/g reads in pUC19). The same
                     ligation was transformed in either JM107 or DH5a. "
BASE COUNT      187 a    135 c    102 g    176 t
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 29;  Length 600;
  Best Local Similarity   87.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGA 23
              ||  ||||||||||||||| |||
Db        188 TAATATCCTCTAGTACTTCACGA 166


RESULT 9
AQ305217/c
LOCUS       AQ305217                608 bp    DNA     linear   GSS 16-DEC-1998
DEFINITION  HS_2019_A2_D07_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2019 Col=14 Row=G, genomic survey
            sequence.
ACCESSION   AQ305217
VERSION     AQ305217.1  GI:4025003
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 608)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2019  row: G  column: 14
            Class: BAC ends
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1. .608
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2019 Col=14 Row=G"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      168 a    150 c    136 g    140 t     14 others
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 28;  Length 608;
  Best Local Similarity   87.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 ATGATCCTCTAGTACTTCTCGAC 24
              |||||||||| ||| ||| ||||
Db        236 ATGATCCTCTGGTAGTTCACGAC 214


RESULT 10
BH022001
LOCUS       BH022001                609 bp    DNA     linear   GSS 29-JUN-2001
DEFINITION  GH_MBb0001P10f Gossypium hirsutum L. Gossypium hirsutum genomic
            clone GH_MBb0001P10f, genomic survey sequence.
ACCESSION   BH022001
VERSION     BH022001.1  GI:14576289
KEYWORDS    GSS.
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Tomkins,J.P., Peterson,D.G., Yang,T.J., Main,D., Wilkins,T.A.,
            Paterson,A.H. and Wing,R.A.
  TITLE     Development of Genomic Resources for Cotton (Gossypium hirsutum
            L.): BAC Library Construction, Preliminary STC Analysis, and
            Identification of Clones Associated With Fiber Development
  JOURNAL   Unpublished
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 4
            High quality sequence stop: 230.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Gossypium hirsutum"
                     /mol_type="genomic DNA"
                     /cultivar="Maxxa"
                     /db_xref="taxon:3635"
                     /clone="GH_MBb0001P10f"
                     /tissue_type="Young leaves"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium hirsutum L."
                     /note="Vector: pCUGIBAC-1; Site_1: HindIII; Site_2: NotI;
                     For more details on library preparation, ordering clones
                     and sequence  analysis see
                     http://www.genome.clemson.edu/projects/stc/cotton/GH_MBb "
BASE COUNT      152 a    156 c     81 g    220 t
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 28;  Length 609;
  Best Local Similarity   87.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGA 23
              || || |||||||||||||||| |
Db        390 TACGAACCTCTAGTACTTCTCTA 412


RESULT 11
CA781659
LOCUS       CA781659                737 bp    mRNA    linear   EST 03-DEC-2002
DEFINITION  031G11AF Infected Arabidopsis Leaf Arabidopsis thaliana cDNA, mRNA
            sequence.
ACCESSION   CA781659
VERSION     CA781659.1  GI:26019692
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
```

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Lundsgaard,M., Emmersen,J., Nielsen,K.L., Wilson,I., Somerville,S.
            and Welinder,K.G.
  TITLE     EST sequencing of Erysiphe cichoracearum infected Arabidopsis
            plants
  JOURNAL   Unpublished
COMMENT     Contact: Karen G. Welinder
            Institut for bioteknologi
            Aalborg Universitet
            Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
            Tel: +45 96358467
            Fax: +45 98141808
            Email: kgw@bio.auc.dk.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /dev_stage="Plant 3 weeks old, three days post infection"
                     /clone_lib="Infected Arabidopsis Leaf"
                     /note="Organ: Leaf; Vector: pBluescript; Mixed cDNA
                     library of Arabidopsis and E. cichoracearum infected leaf
                     from three weeks old Arabidopsis plants. Plants were
                     harvested 3 days after infection and mRNA oligo dT
                     selected."
BASE COUNT      204 a    137 c    156 g    240 t
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 14;  Length 737;
  Best Local Similarity   87.0%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGA 23
              ||||||| || ||||||||| ||
Db        668 TATGATCATCAAGTACTTCTTGA 690


RESULT 12
BZ506781
LOCUS       BZ506781                741 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  BONFA77TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONFA77,
            genomic survey sequence.
ACCESSION   BZ506781
VERSION     BZ506781.1  GI:27028149
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished
COMMENT     Other_GSSs: BONFA77TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONFA77"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      273 a    122 c    115 g    231 t
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 29;  Length 741;
  Best Local Similarity   87.0%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGA 23
              ||||||| |||||||| | ||||
Db        276 TATGATCATCTAGTACATTTCGA 298


RESULT 13
BZ998034
LOCUS       BZ998034                799 bp    DNA     linear   GSS 25-MAR-2003
DEFINITION  PUGFY17TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa375D09,
            genomic survey sequence.
ACCESSION   BZ998034
VERSION     BZ998034.1  GI:29241451
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Other_GSSs: PUGFY17TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .799
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa375D09"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      215 a    142 c    181 g    261 t
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 29;  Length 799;
  Best Local Similarity   87.0%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGA 23
              ||  |||||||||||||| |||
Db        528 TAATATCCTCTAGTACTTCACGA 550


RESULT 14
AQ862478
LOCUS       AQ862478                815 bp    DNA     linear   GSS 03-NOV-1999
DEFINITION  nbeb0019A21f CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica
            cultivar-group) genomic clone nbeb0019A21f, genomic survey
            sequence.
ACCESSION   AQ862478
```

```
VERSION     AQ862478.1  GI:6212935
KEYWORDS    GSS.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 815)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 46
            High quality sequence stop: 279.
FEATURES             Location/Qualifiers
     source          1. .815
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="nbeb0019A21f"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT      215 a    213 c     84 g    298 t      5 others
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 28;  Length 815;
  Best Local Similarity   87.0%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGA 23
              ||| |||||||| |||||||| |
Db        789 TATTATCCTCTAATACTTCTCTA 811


RESULT 15
BZ576392
LOCUS       BZ576392               1661 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_4912.y2 msh Pseudomonas aeruginosa genomic clone msh2_4912,
            genomic survey sequence.
ACCESSION   BZ576392
VERSION     BZ576392.1  GI:27211453
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1661)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1661
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_4912"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
BASE COUNT      340 a    434 c    265 g    618 t      4 others
ORIGIN

  Query Match             75.8%;  Score 18.2;  DB 29;  Length 1661;
  Best Local Similarity   87.0%;  Pred. No. 1.9e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 ATGATCCTCTAGTACTTCTCGAC 24
              || |||||||  ||||||||||||
Db        656 ATCATCCTCTCCTACTTCTCGAC 678


Search completed: February 16, 2004, 13:40:58
Job time : 97.2277 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:56:25 ; Search time 114.587 Seconds
(without alignments)
8568.399 Million cell updates/sec

Title: US-10-676-079-5
Perfect score: 24
Sequence: 1 tatgatcctctagtacttctcgac 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		24	100.0	24	6	AR125606	AR125606 Sequence
2		24	100.0	24	6	AR194192	AR194192 Sequence
3		24	100.0	24	6	AR221288	AR221288 Sequence
4		24	100.0	24	6	AR243206	AR243206 Sequence
5		24	100.0	24	6	AR287438	AR287438 Sequence
6	c	24	100.0	604	6	AX044389	AX044389 Sequence
7	c	24	100.0	860	6	AX044388	AX044388 Sequence
8	c	24	100.0	956	6	AX044387	AX044387 Sequence
9	c	24	100.0	1430	6	AX188665	AX188665 Sequence
10	c	24	100.0	2062	6	AX044385	AX044385 Sequence
11	c	24	100.0	2062	6	AX044386	AX044386 Sequence
12	c	24	100.0	2436	6	AX044392	AX044392 Sequence
13	c	24	100.0	2436	6	AX044395	AX044395 Sequence
14	c	24	100.0	2692	6	AX044391	AX044391 Sequence
15	c	24	100.0	2692	6	AX044394	AX044394 Sequence
16	c	24	100.0	2788	6	AX044390	AX044390 Sequence
17	c	24	100.0	2788	6	AX044393	AX044393 Sequence
18	c	24	100.0	7668	6	BD170085	BD170085 Novel ADA
19	c	24	100.0	8435	6	BD170083	BD170083 Novel ADA
20	c	24	100.0	8505	6	BD170084	BD170084 Novel ADA
21	c	23	95.8	934	6	AX077879	AX077879 Sequence
22	c	18.8	78.3	164000	9	AL591682	AL591682 Human DNA
23		18.4	76.7	110000	2	LMFLCHR36_22	Continuation (23 o
24		18.4	76.7	169529	2	AC131586	AC131586 Mus muscu
25	c	18.2	75.8	404	8	ATH529465	AJ529465 Arabidops
26		18.2	75.8	1342	8	AY142502	AY142502 Arabidops
27	c	18.2	75.8	2000	6	AX461217	AX461217 Sequence
28		18.2	75.8	2338	8	BT005733	BT005733 Arabidops
29		18.2	75.8	3006	2	AC018026	AC018026 Drosophil
30	c	18.2	75.8	24942	8	YSCD9819	U51031 Saccharomyc
31		18.2	75.8	45980	8	ATT24A18	AL035680 Arabidops
32		18.2	75.8	83513	8	ATT20E23	AL133363 Arabidops
33	c	18.2	75.8	105549	10	AL844906	AL844906 Mouse DNA
34		18.2	75.8	108409	14	AF083424	AF083424 Ateline h
35	c	18.2	75.8	153125	9	AC104456	AC104456 Homo sapi
36		18.2	75.8	165617	5	AL929502	AL929502 Zebrafish
37	c	18.2	75.8	171949	9	AC008652	AC008652 Homo sapi
38	c	18.2	75.8	173810	10	AC083815	AC083815 Mus muscu
39		18.2	75.8	198372	8	ATCHRIV66	AL161566 Arabidops
40	c	18.2	75.8	207639	10	AL844581	AL844581 Mouse DNA
41		18.2	75.8	209114	9	AC008383	AC008383 Homo sapi
42		18.2	75.8	230128	10	AC098707	AC098707 Mus muscu
43		18.2	75.8	235996	2	AC096147	AC096147 Rattus no
44	c	18.2	75.8	242679	2	AC117867	AC117867 Rattus no
45		18.2	75.8	279353	2	AC120587	AC120587 Rattus no

ALIGNMENTS

RESULT 1
AR125606
LOCUS AR125606 24 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 5 from patent US 6177545.
ACCESSION AR125606
VERSION AR125606.1 GI:14111668
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
TITLE Heparanase specific molecular probes and their use in research and medical applications
JOURNAL Patent: US 6177545-A 5 23-JAN-2001;

```
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        5 a      7 c      3 g      9 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.56;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db          1 TATGATCCTCTAGTACTTCTCGAC 24


RESULT 2
AR194192
LOCUS       AR194192                 24 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 5 from patent US 6348344.
ACCESSION   AR194192
VERSION     AR194192.1  GI:20240784
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M.,
            Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Schlomi,Y.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6348344-A 5 19-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        5 a      7 c      3 g      9 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.56;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db          1 TATGATCCTCTAGTACTTCTCGAC 24


RESULT 3
AR221288
LOCUS       AR221288                 24 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 5 from patent US 6426209.
ACCESSION   AR221288
VERSION     AR221288.1  GI:23328259
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Ayal-Hershkovitz,M., Pecker,I. and Yacoby-Zeevi,O.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6426209-A 5 30-JUL-2002;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        5 a      7 c      3 g      9 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.56;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db          1 TATGATCCTCTAGTACTTCTCGAC 24


RESULT 4
AR243206
LOCUS       AR243206                 24 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 5 from patent US 6475763.
ACCESSION   AR243206
VERSION     AR243206.1  GI:27290321
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M.,
            Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Shlomi,Y.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6475763-A 5 05-NOV-2002;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        5 a      7 c      3 g      9 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.56;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db          1 TATGATCCTCTAGTACTTCTCGAC 24


RESULT 5
AR287438
LOCUS       AR287438                 24 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 5 from patent US 6531129.
ACCESSION   AR287438
VERSION     AR287438.1  GI:29725132
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6531129-A 5 11-MAR-2003;
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unknown"
BASE COUNT        5 a      7 c      3 g      9 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 0.56;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db          1 TATGATCCTCTAGTACTTCTCGAC 24


RESULT 6
AX044389/c
LOCUS       AX044389                 604 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 5 from Patent WO0066755.
ACCESSION   AX044389
VERSION     AX044389.1  GI:11343267
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

```
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 5 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of RTB3 in pFASTBAC1"
BASE COUNT      180 a    121 c    134 g    169 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 604;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db        599 TATGATCCTCTAGTACTTCTCGAC 576


RESULT 7
AX044388/c
LOCUS       AX044388                 860 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 4 from Patent WO0066755.
ACCESSION   AX044388
VERSION     AX044388.1  GI:11343266
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 4 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .860
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of RTB2 in pFASTBAC1"
BASE COUNT      255 a    165 c    200 g    240 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 860;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db        855 TATGATCCTCTAGTACTTCTCGAC 832


RESULT 8
AX044387/c
LOCUS       AX044387                 956 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 3 from Patent WO0066755.
ACCESSION   AX044387
VERSION     AX044387.1  GI:11343265
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 3 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of RTB1 in pFASTBAC1"
BASE COUNT      285 a    189 c    215 g    267 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 956;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db        951 TATGATCCTCTAGTACTTCTCGAC 928


RESULT 9
AX188665/c
LOCUS       AX188665                1430 bp    DNA     linear   PAT 10-AUG-2001
DEFINITION  Sequence 13 from Patent WO0147963.
ACCESSION   AX188665
VERSION     AX188665.1  GI:15142264
KEYWORDS    .
SOURCE      Placobdella papillifera
  ORGANISM  Placobdella papillifera
            Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
            Rhynchobdellida; Glossiphoniidae; Placobdella.
REFERENCE   1
  AUTHORS   Finney,S., Seale,L. and Wallis,R.B.
  TITLE     Inhibitors of complement activation, their preparation and use
  JOURNAL   Patent: WO 0147963-A 13 05-JUL-2001;
            Bio-Discovery Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .1430
                     /organism="Placobdella papillifera"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:168682"
BASE COUNT      419 a    296 c    296 g    419 t
ORIGIN

  Query Match            100.0%;  Score 24;  DB 6;  Length 1430;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       1126 TATGATCCTCTAGTACTTCTCGAC 1103


RESULT 10
AX044385/c
LOCUS       AX044385                2062 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 1 from Patent WO0066755.
ACCESSION   AX044385
VERSION     AX044385.1  GI:11343263
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 1 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .2062
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of CryIA(b) in pFASTBAC1"
BASE COUNT      538 a    552 c    440 g    532 t
ORIGIN
```

```
  Query Match             100.0%;  Score 24;  DB 6;  Length 2062;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2057 TATGATCCTCTAGTACTTCTCGAC 2034


RESULT 11
AX044386/c
LOCUS       AX044386               2062 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 2 from Patent WO0066755.
ACCESSION   AX044386
VERSION     AX044386.1  GI:11343264
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 2 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .2062
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of CryIA(c) in pFASTBAC1"
BASE COUNT      544 a    532 c    446 g    540 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 2062;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2057 TATGATCCTCTAGTACTTCTCGAC 2034


RESULT 12
AX044392/c
LOCUS       AX044392               2436 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 8 from Patent WO0066755.
ACCESSION   AX044392
VERSION     AX044392.1  GI:11343270
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 8 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .2436
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of CryIA(b)-RTB3 in pFASTBAC1"
BASE COUNT      654 a    617 c    521 g    644 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 2436;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2431 TATGATCCTCTAGTACTTCTCGAC 2408


RESULT 13
AX044395/c
LOCUS       AX044395               2436 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 11 from Patent WO0066755.
ACCESSION   AX044395
VERSION     AX044395.1  GI:11343273
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 11 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .2436
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of CryIA(c)-RTB3 in pFASTBAC1"
BASE COUNT      660 a    597 c    527 g    652 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 2436;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2431 TATGATCCTCTAGTACTTCTCGAC 2408


RESULT 14
AX044391/c
LOCUS       AX044391               2692 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 7 from Patent WO0066755.
ACCESSION   AX044391
VERSION     AX044391.1  GI:11343269
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 7 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .2692
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of CryIA(b)-RTB2 in pFASTBAC1"
BASE COUNT      730 a    660 c    586 g    716 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 2692;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2687 TATGATCCTCTAGTACTTCTCGAC 2664


RESULT 15
AX044394/c
LOCUS       AX044394               2692 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 10 from Patent WO0066755.
```

```
ACCESSION   AX044394
VERSION     AX044394.1  GI:11343272
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Christou,P. and Mehlo,L.
  TITLE     Pesticidal fusions
  JOURNAL   Patent: WO 0066755-A 10 09-NOV-2000;
            Plant Bioscience Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .2692
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence of CryIA(c)-RTB2 in pFASTBAC1"
BASE COUNT      736 a    640 c    592 g    724 t
ORIGIN

  Query Match             100.0%;  Score 24;  DB 6;  Length 2692;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGATCCTCTAGTACTTCTCGAC 24
              ||||||||||||||||||||||||
Db       2687 TATGATCCTCTAGTACTTCTCGAC 2664


Search completed: February 16, 2004, 11:43:02
Job time : 117.587 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:55:00 ; Search time 12.7569 Seconds
(without alignments)
5501.769 Million cell updates/sec

Title: US-10-676-079-4
Perfect score: 26
Sequence: 1 cgcatatgcaggacgtcgtggacctg 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	100.0	26	21	AAZ39196	Human heparanase P
2	26	100.0	26	21	AAZ33291	Human heparanase P
3	20	76.9	30	24	ABN86005	Human heparanase g
4	19.6	75.4	474	22	AAF93984	Primer specific fo
5	19.6	75.4	605	24	ABN86000	Human heparanase g
6	19.6	75.4	1419	24	ABN86003	Human heparanase u
7	19.6	75.4	1593	20	AAZ11236	Human pre-prohepar
8	19.6	75.4	1713	20	AAX37259	Human heparanase e
9	19.6	75.4	1721	20	AAX35648	cDNA encoding a hu
10	19.6	75.4	1721	21	AAA75051	cDNA encoding a hu
11	19.6	75.4	1721	21	AAZ39195	Human heparanase e
12	19.6	75.4	1721	21	AAZ33290	Human heparanase n
13	19.6	75.4	1721	22	AAA91112	Human heparanase,
14	19.6	75.4	1722	22	AAF93788	Human cDNA encodin
15	19.6	75.4	1723	20	AAX37260	Seq ID No: 14 of W
16	19.6	75.4	1724	22	AAH20940	Human heparanase i
17	19.6	75.4	1899	20	AAX35650	cDNA encoding a hu
18	19.6	75.4	1899	21	AAA75053	cDNA encoding a hu
19	19.6	75.4	3726	20	AAX86671	cDNA encoding a hu
20	19.6	75.4	44848	21	AAA75080	Nucleotide sequenc
c 21	19.2	73.8	444	24	ABN90554	Staphylococcus epi
c 22	19.2	73.8	1368	24	ABN90553	Staphylococcus epi
23	19	73.1	1669	25	ABZ22816	Human heparanase e
24	18.6	71.5	35	21	AAZ39200	Human heparanase P
c 25	18.6	71.5	2588	25	ABT16043	NOVX related polyn
c 26	18.6	71.5	4345	22	AAF61716	Human CTAGE-2 enco
27	18.6	71.5	11854	20	AAX13243	Enterococcus faeca
28	18.6	71.5	11854	24	ABS99038	Enterococcus faeca
c 29	18.4	70.8	5770	23	AAS84301	DNA encoding novel
c 30	18.2	70.0	1298	24	AAL42627	Coffee theobromine
c 31	18.2	70.0	1302	24	AAK99760	1302nt DNA encodin
c 32	18.2	70.0	1302	24	AAK99773	1302nt RNA encodin
c 33	18.2	70.0	1304	24	AAL42629	Coffee theobromine
c 34	18.2	70.0	1360	24	AAL42628	Coffee theobromine
35	18	69.2	33	19	AAV71120	PCR primer used to
36	18	69.2	33	20	AAZ30938	Thermus thermophil
37	18	69.2	60	24	ABN40291	Human spliced tran
c 38	18	69.2	177	21	AAA36803	Human dysferlin re
c 39	18	69.2	177	21	AAX82894	Human dysferlin DN
40	18	69.2	262	25	ABX31423	Human GDP-mannose
41	18	69.2	1584	24	ABL40753	Chicken signal pep
42	18	69.2	2032	24	ABL92126	Human Tumour Endot
43	18	69.2	2032	25	ABX72051	DNA encoding human
44	18	69.2	2312	25	ABZ09890	Human 5' and/or re
c 45	18	69.2	2668	24	ABQ70896	Listeria monocytog

ALIGNMENTS

RESULT 1
AAZ39196
ID AAZ39196 standard; DNA; 26 BP.
XX
AC AAZ39196;
XX
DT 02-MAR-2000 (first entry)
XX
DE Human heparanase PCR sense primer SEQ ID NO:4.
XX
KW Human; heparanase; hpa; genetic modification; expression; anticancer;
KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW micrometastasis; autoimmune lesion; kidney failure; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9957244-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09256.
XX
PR 01-MAY-1998; 98US-0071618.
PR 02-MAR-1999; 99US-0260038.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (FRIE/) FRIEDMAN M M.
XX
PI Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I, Peleg Y;
PI Shlomi Y;
XX
DR WPI; 2000-062144/05.
XX
PT Engineered cells that express recombinant heparanase, useful
PT therapeutically, e.g. for treating angiogenesis and to screen for
PT specific inhibitors, potential anticancer agents -
XX
PS Example 1; Page 36; 118pp; English.
XX
CC The present invention describes genetically modified cells (A) containing
CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,
CC and express recombinant heparanase (II). Heparanase cleaves heparan
CC sulphate (HS) at specific intrachain sites, resulting in release of
CC heparin-binding growth factors, enzymes and proteins that are sequestered
CC by HS in basement membranes, extracellular matrix or cell surfaces. It
CC may also be implicated in tumour angiogenesis and metastases. (II) is
CC potentially useful in wound healing and for treating angiogenesis,
CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
CC infection and cystic fibrosis. It can also be used to neutralise heparin
CC (an alternative to protamine) and to screen for specific inhibitors
CC (potentially useful for treating cancer and metastases). Antibodies
CC raised against (II) are used for immunodetection and diagnosis of
CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
CC in large quantities, in a form that is homogeneously processed and
CC activated/neutralised by a dedicated protease. The present sequence
CC represents a PCR primer for human heparanase, which is used in an
CC example from the present invention.
XX
SQ Sequence 26 BP; 5 A; 7 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 21; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCATATGCAGGACGTCGTGGACCTG 26
||||||||||||||||||||||||||
Db 1 CGCATATGCAGGACGTCGTGGACCTG 26

RESULT 2
AAZ33291
ID AAZ33291 standard; DNA; 26 BP.
XX
AC AAZ33291;
XX
DT 21-FEB-2000 (first entry)
XX
DE Human heparanase PCR sense primer SEQ ID NO:4.
XX
KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW autoimmune disease; anticancer; kidney disease; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09255.
XX
PR 01-MAY-1998; 98US-0071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
XX
PT Heparanase-specific molecular probes useful for diagnosis and
PT treatment, e.g. of tumors, and for targeted drug delivery -
XX
PS Example; Page 27; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and
CC anti-heparanase antibodies are used to detect or quantify the expression
CC of heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary,
CC cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC syndrome, sepsis and inflammatory or autoimmune disease), for targeted
CC drug delivery (e.g. of anticancer agents) and as research reagents.
CC The present sequence represents a PCR primer for human heparanase, which
CC is used in an example from the present invention for the construction of
CC a heparanase expression vector.
XX
SQ Sequence 26 BP; 5 A; 7 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 21; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCATATGCAGGACGTCGTGGACCTG 26
||||||||||||||||||||||||||
Db 1 CGCATATGCAGGACGTCGTGGACCTG 26

RESULT 3
ABN86005
ID ABN86005 standard; DNA; 30 BP.
XX
AC ABN86005;
XX
DT 06-SEP-2002 (first entry)
XX
DE Human heparanase gene specific primer HP-6.
XX
KW Human; heparanase; cytostatic; vasotropic; antidiabetic; anti-HIV;
KW ophthalmological; antirheumatic; antiarthritic; antipsoriatic;
KW antianaemic; neuroprotective; nootropic; cerebroprotective;
KW antibacterial; virucide; protozoacide; fungicide; antiinflammatory;
KW cardiant; immunosuppressive; tumour metastasis; inflammatory disease;
KW allograft rejection; cell migration; angiogenesis; basement membrane;
KW extracellular matrix; cancer; ischaemia; diabetic retinopathy;
KW macular degeneration; rheumatoid arthritis; psoriasis; HIV infection;
KW sickle cell anaemia; Alzheimer's disease; muscular dystrophy;
KW neurodegenerative disease; vascular disease; cardiovascular disease;
KW cystic fibrosis; stroke; gene therapy; PCR; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200244353-A2.
XX
PD 06-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US44798.
XX
PR 30-NOV-2000; 2000US-250690P.
XX
PA (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI Wolffe AP, Qi H;

```
XX
DR   WPI; 2002-527708/56.
XX
PT   New heparanase polynucleotide, useful for controlling disease states
PT   such as tumour metastasis, inflammatory diseases and allograft rejection
PT    -
XX
PS   Example 1; Page 44; 72pp; English.
XX
CC   The invention relates to novel heparanase sequences, particularly novel
CC   sequences from the regulatory regions upstream and downstream of the
CC   coding region. The activity of polynucleotides of the invention may be
CC   described as, cytostatic, vasotropic, antidiabetic, anti-HIV,
CC   ophthalmological, antirheumatic, antiarthritic, antipsoriatic,
CC   antianaemic, neuroprotective, nootropic, cerebroprotective,
CC   antibacterial, virucide, protozoacide, fungicide, antiinflammatory,
CC   cardiant and immunosuppressive. Modulating expression of heparanase gene
CC   using constructs of the invention is useful for facilitating targeted
CC   control of disease states such as tumour metastasis, inflammatory
CC   diseases, allograft rejection, and for inhibiting processes such as cell
CC   migration, angiogenesis, and degradation of the basement membrane and/or
CC   extracellular matrix. Heparanase-targeted DNA binding domains modulates
CC   gene expression, and are useful for therapeutic or prophylactic
CC   applications, for e.g. cancer, ischaemia, diabetic retinopathy, macular
CC   degeneration, rheumatoid arthritis, psoriasis, HIV infection, sickle cell
CC   anaemia, Alzheimer's disease, muscular dystrophy, neurodegenerative
CC   diseases, vascular disease, cardiovascular disease, cystic fibrosis,
CC   stroke, and bacterial, protozoal, fungal and viral infection. Constructs
CC   of the invention may also be useful in gene therapy. The current sequence
CC   represents a human heparanase gene specific primer designated HP-6. This
CC   was used in the determination of nucleotide sequences in the human
CC   heparanase gene and flanking regions.
XX
SQ   Sequence 30 BP; 4 A; 9 C; 9 G; 8 T; 0 other;

  Query Match             76.9%;  Score 20;  DB 24;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 TGCAGGACGTCGTGGACCTG 26
              ||||||||||||||||||||
Db          2 TGCAGGACGTCGTGGACCTG 21


RESULT 4
AAF93984
ID   AAF93984 standard; DNA; 474 BP.
XX
AC   AAF93984;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Primer specific for DNA encoding secretory/membrane protein SEQ ID 418.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS   Synthetic.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 4; SEQ ID 418; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 474 BP; 86 A; 154 C; 127 G; 101 T; 6 other;

  Query Match             75.4%;  Score 19.6;  DB 22;  Length 474;
  Best Local Similarity   84.6%;  Pred. No. 38;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        161 CGCAAGCACAGGACGTCGTGGACCTG 186


RESULT 5
ABN86000
ID   ABN86000 standard; DNA; 605 BP.
XX
AC   ABN86000;
XX
DT   06-SEP-2002  (first entry)
XX
DE   Human heparanase gene fragment.
XX
KW   Human; heparanase; cytostatic; vasotropic; antidiabetic; anti-HIV;
KW   ophthalmological; antirheumatic; antiarthritic; antipsoriatic;
KW   antianaemic; neuroprotective; nootropic; cerebroprotective;
KW   antibacterial; virucide; protozoacide; fungicide;  antiinflammatory;
KW   cardiant; immunosuppressive; tumour metastasis; inflammatory disease;
KW   allograft rejection; cell migration; angiogenesis; basement membrane;
KW   extracellular matrix; cancer; ischaemia; diabetic retinopathy;
KW   macular degeneration; rheumatoid arthritis; psoriasis; HIV infection;
KW   sickle cell anaemia; Alzheimer's disease; muscular dystrophy;
KW   neurodegenerative disease; vascular disease; cardiovascular disease;
KW   cystic fibrosis; stroke; gene therapy; chromosome 4; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..605
FT                   /*tag=  a
FT                   /product= "heparanase protein fragment"
FT                   /note= "contains 2 introns; one full intron and 1 intron
FT                   fragment"
```

```
FT   exon            1..105
FT                   /*tag=  b
FT                   /number= 1
FT   intron          106..339
FT                   /*tag=  c
FT                   /number= 1
FT   exon            340..602
FT                   /*tag=  d
FT                   /number= 2
FT   misc_signal     371..373
FT                   /*tag=  e
FT                   /note= "start codon"
FT   intron          603..605
FT                   /*tag=  f
FT                   /number= 2
FT                   /note= "a small fragment only of this intron is given"
XX
PN   WO200244353-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US44798.
XX
PR   30-NOV-2000; 2000US-250690P.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Wolffe AP,  Qi H;
XX
DR   WPI; 2002-527708/56.
XX
PT   New heparanase polynucleotide, useful for controlling disease states
PT   such as tumour metastasis, inflammatory diseases and allograft rejection
PT    -
XX
PS   Example 1; Fig 1; 72pp; English.
XX
CC   The invention relates to novel heparanase sequences, particularly novel
CC   sequences from the regulatory regions upstream and downstream of the
CC   coding region. The activity of polynucleotides of the invention may be
CC   described as, cytostatic, vasotropic, antidiabetic, anti-HIV,
CC   ophthalmological, antirheumatic, antiarthritic, antipsoriatic,
CC   antianaemic, neuroprotective, nootropic, cerebroprotective,
CC   antibacterial, virucide, protozoacide, fungicide, antiinflammatory,
CC   cardiant and immunosuppressive. Modulating expression of heparanase gene
CC   using constructs of the invention is useful for facilitating targeted
CC   control of disease states such as tumour metastasis, inflammatory
CC   diseases, allograft rejection, and for inhibiting processes such as cell
CC   migration, angiogenesis, and degradation of the basement membrane and/or
CC   extracellular matrix. Heparanase-targeted DNA binding domains modulates
CC   gene expression, and are useful for therapeutic or prophylactic
CC   applications, for e.g. cancer, ischaemia, diabetic retinopathy, macular
CC   degeneration, rheumatoid arthritis, psoriasis, HIV infection, sickle cell
CC   anaemia, Alzheimer's disease, muscular dystrophy, neurodegenerative
CC   diseases, vascular disease, cardiovascular disease, cystic fibrosis,
CC   stroke, and bacterial, protozoal, fungal and viral infection. Constructs
CC   of the invention may also be useful in gene therapy. The current sequence
CC   represents a human heparanase gene fragment which is located on
CC   chromosome 4.
XX
SQ   Sequence 605 BP; 96 A; 187 C; 230 G; 92 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 24;  Length 605;
  Best Local Similarity   84.6%;  Pred. No. 38;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        468 CGCAAGCACAGGACGTCGTGGACCTG 493


RESULT 6
ABN86003
ID   ABN86003 standard; DNA; 1419 BP.
XX
AC   ABN86003;
XX
DT   06-SEP-2002  (first entry)
XX
DE   Human heparanase upstream sequence containing exons 1 and 2.
XX
KW   Human; heparanase; cytostatic; vasotropic; antidiabetic; anti-HIV;
KW   ophthalmological; antirheumatic; antiarthritic; antipsoriatic;
KW   antianaemic; neuroprotective; nootropic; cerebroprotective;
KW   antibacterial; virucide; protozoacide; fungicide; antiinflammatory;
KW   cardiant; immunosuppressive; tumour metastasis; inflammatory disease;
KW   allograft rejection; cell migration; angiogenesis; basement membrane;
KW   extracellular matrix; cancer; ischaemia; diabetic retinopathy;
KW   macular degeneration; rheumatoid arthritis; psoriasis; HIV infection;
KW   sickle cell anaemia; Alzheimer's disease; muscular dystrophy;
KW   neurodegenerative disease; vascular disease; cardiovascular disease;
KW   cystic fibrosis; stroke; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200244353-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US44798.
XX
PR   30-NOV-2000; 2000US-250690P.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Wolffe AP,  Qi H;
XX
DR   WPI; 2002-527708/56.
XX
PT   New heparanase polynucleotide, useful for controlling disease states
PT   such as tumour metastasis, inflammatory diseases and allograft rejection
PT    -
XX
PS   Example 1; Fig 4; 72pp; English.
XX
CC   The invention relates to novel heparanase sequences, particularly novel
CC   sequences from the regulatory regions upstream and downstream of the
CC   coding region. The activity of polynucleotides of the invention may be
CC   described as, cytostatic, vasotropic, antidiabetic, anti-HIV,
CC   ophthalmological, antirheumatic, antiarthritic, antipsoriatic,
CC   antianaemic, neuroprotective, nootropic, cerebroprotective,
CC   antibacterial, virucide, protozoacide, fungicide, antiinflammatory,
CC   cardiant and immunosuppressive. Modulating expression of heparanase gene
CC   using constructs of the invention is useful for facilitating targeted
CC   control of disease states such as tumour metastasis, inflammatory
CC   diseases, allograft rejection, and for inhibiting processes such as cell
CC   migration, angiogenesis, and degradation of the basement membrane and/or
CC   extracellular matrix. Heparanase-targeted DNA binding domains modulates
CC   gene expression, and are useful for therapeutic or prophylactic
CC   applications, for e.g. cancer, ischaemia, diabetic retinopathy, macular
CC   degeneration, rheumatoid arthritis, psoriasis, HIV infection, sickle cell
CC   anaemia, Alzheimer's disease, muscular dystrophy, neurodegenerative
CC   diseases, vascular disease, cardiovascular disease, cystic fibrosis,
CC   stroke, and bacterial, protozoal, fungal and viral infection. Constructs
CC   of the invention may also be useful in gene therapy. The current sequence
CC   represents the human heparanase upstream sequence containing exons 1 and
CC   2.
XX
SQ   Sequence 1419 BP; 329 A; 352 C; 452 G; 286 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 24;  Length 1419;
  Best Local Similarity   84.6%;  Pred. No. 41;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
```

```
Db      1282 CGCAAGCACAGGACGTCGTGGACCTG 1307

RESULT 7
AAZ11236
ID   AAZ11236 standard; cDNA; 1593 BP.
XX
AC   AAZ11236;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Human pre-proheparanase coding sequence.
XX
KW   Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
KW   inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
KW   heparin degradation; anticoagulant neutralisation; asthma; CNS disease;
KW   inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
KW   tumour growth; fibroproliferative disorder; neurodegenerative disease;
KW   therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1593
FT                   /*tag=  a
FT                   /product= pre-proheparanase
XX
PN   WO9943830-A2.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999;   99WO-US01489.
XX
PR   26-MAR-1998;   98US-0079401.
PR   24-FEB-1998;   98US-0075706.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Fairbanks MB,  Heinrikson RL,  Mildner AM;
XX
DR   WPI; 1999-540598/45.
DR   P-PSDB; AAY34173.
XX
PT   New isolated platelet heparanase polypeptides, used to develop
PT   products for, e.g. wound healing and blocking angiogenesis
XX
PS   Claim 2; Fig 7; 57pp; English.
XX
CC   This sequence encodes the human pre-proheparanase of the invention. This
CC   sequence was isolated from human platelets. The heparanase can be used
CC   for identifying agents which alter heparanase activity. The heparanase
CC   can be used for wound healing or for blocking angiogenesis or
CC   inflammation. It can be used for treating e.g. psoriasis, diabetic
CC   retinopathy or solid tumours, or for the degradation of heparin and the
CC   neutralisation of heparin's anticoagulant properties during surgery.
CC   Inhibitors of heparanase activity can be used in the treatment of
CC   arthritis, asthma, and other inflammatory diseases, vascular restenosis,
CC   atherosclerosis, tumour growth and progression, fibroproliferative
CC   disorders, and central nervous system (CNS) and neurodegenerative
CC   diseases. The products can also be used for detection and diagnosis. The
CC   purified heparanase, both recombinantly produced human heparanase and
CC   heparanase isolated from human platelet activity, allows for the
CC   convenient selection of compounds having anti-heparanase activity,
CC   i.e. inhibitors of heparanase activity, by measuring inhibition of
CC   heparanase activity. Inhibition of heparanase activity can be measured by
CC   blocking heparanase-mediated release of radioactive fragments from in
CC   vivo radiolabelled (HSPG)/heparin.
XX
SQ   Sequence 1593 BP; 426 A; 370 C; 369 G; 428 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 20;  Length 1593;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db         59 CGCAAGCACAGGACGTCGTGGACCTG 84


RESULT 8
AAX37259
ID   AAX37259 standard; DNA; 1713 BP.
XX
AC   AAX37259;
XX
DT   21-JUL-1999  (first entry)
XX
DE   Human heparanase enzyme encoding DNA.
XX
KW   Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW   metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW   arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW   human; HSPG; ss.
XX
OS   Homo sapiens.
XX
PN   WO9921975-A1.
XX
PD   06-MAY-1999.
XX
PF   28-OCT-1998;   98WO-AU00898.
XX
PR   09-DEC-1997;   97AU-0000812.
PR   28-OCT-1997;   97AU-0000062.
XX
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Freeman CG,  Hamdorf BJ,  Hulett MD,  Parish CR;
XX
DR   WPI; 1999-312956/26.
DR   P-PSDB; AAY17082.
XX
PT   Polynucleotides encoding mammalian endoglucuronidases, especially
PT   heparanases, useful to promote wound healing
XX
PS   Claim 3; Page 69-73; 112pp; English.
XX
CC   The invention relates to nucleic acid sequences that encode heparanase
CC   enzymes having  endoglucuronidase activity. Recombinant heparanases are
CC   capable of removing the HS side chain from heparan sulfate proteoglycan
CC   (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC   inhibit heparanase, this is useful for treatment of a physiological or
CC   medical condition associated with elevated heparanase activity, such as
CC   metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC   arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC   rat heparanases can be used to enhance wound healing, especially
CC   associated with tissue development and repair. The conditions mentioned
CC   above can be diagnosed using specific antibodies, and also using primers
CC   and probes specific for the heparanase polynucleotides. Other uses of the
CC   heparanases include sequencing sulfated molecules such as HSPG. The
CC   present sequence represents a DNA encoding human heparanase.
XX
SQ   Sequence 1713 BP; 460 A; 404 C; 406 G; 443 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 20;  Length 1713;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        143 CGCAAGCACAGGACGTCGTGGACCTG 168


RESULT 9
AAX35648
ID   AAX35648 standard; cDNA; 1721 BP.
```

```
XX
AC   AAX35648;
XX
DT   09-JUL-1999  (first entry)
XX
DE   cDNA encoding a human heparanase protein.
XX
KW   Heparanase; hpa; modulator; heparin-binding growth factor;
KW   cellular response; cytokine; cell interaction; plasma lipoprotein;
KW   cellular susceptibility; infection; disintegration;
KW   neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW   athersclerosis; inflammation; neurodegenerative disease; neutralise;
KW   plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO9911798-A1.
XX
PD   11-MAR-1999.
XX
PF   31-AUG-1998;   98WO-US17954.
XX
PR   02-JUL-1998;   98US-0109386.
PR   02-SEP-1997;   97US-0922170.
XX
PA   (FRIE/) FRIEDMAN M M.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI   Feinstein E,  Pecker I,  Vlodavsky I;
XX
DR   WPI; 1999-302255/25.
DR   P-PSDB; AAY02345.
XX
PT   New human polynucleotide useful for treating angiogenesis,
PT   restenosis, and inflammation
XX
PS   Claim 4; Fig 1; 63pp; English.
XX
CC   The specification describes a polypeptide having heparanase (hpa)
CC   activity. The recombinant protein is used as a modulator of
CC   heparin-binding growth factors, cellular responses to heparin-binding
CC   growth factors and cytokines, cell interaction with plasma lipoproteins,
CC   cellular susceptibility to viral, protozoal and bacterial infections
CC   or disintegration of neurodegenerative plaques. Heparanase may be
CC   useful for conditions such as wound healing, angiogenesis, restenosis,
CC   athersclerosis, inflammation, neurodegenerative diseases, and viral
CC   infections. Mammalian heparanase can be used to neutralize plasma
CC   heparin, and anti-heparanase antibodies may be applied for
CC   immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC   and renal failure in biopsy specimens, plasma samples, and body fluids.
CC   The present sequence encodes human heparanase.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 20;  Length 1721;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        160 CGCAAGCACAGGACGTCGTGGACCTG 185


RESULT 10
AAA75051
ID   AAA75051 standard; cDNA; 1721 BP.
XX
AC   AAA75051;
XX
DT   16-JAN-2001  (first entry)
XX
DE   cDNA encoding a human heparanase polypeptide.
XX
KW   Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW   heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW   wound healing; infection; burn; angiogenesis; restenosis;
KW   atherosclerosis; inflammation; neurodegenerative disease;
KW   Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1693
FT                   /*tag=  a
FT                   /product= "heparanase"
FT   stem_loop       698..724
FT                   /*tag=  b
FT                   /note= "these nucleotides are likely to be involved
FT                          in forming stem and loop structures"
XX
PN   WO200052178-A1.
XX
PD   08-SEP-2000.
XX
PF   14-FEB-2000; 2000WO-US03542.
XX
PR   01-MAR-1999;   99US-0258892.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Feinstein E;
XX
DR   WPI; 2000-579289/54.
DR   P-PSDB; AAB08849.
XX
PT   New polynucleotides encoding a polypeptide having heparanase activity,
PT   useful in wound healing and in gene therapy, particularly in treating
PT   tumour, inflammation, autoimmunity, neurodegenerative diseases  -
XX
PS   Claim 9; Fig 1; 152pp; English.
XX
CC   The present sequence encodes a human protein with heparanase catalytic
CC   activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC   particularly in treating tumour, inflammation or autoimmunity.
CC   Particularly, the polynucleotide is useful in modulating the
CC   bioavailability of heparin-binding growth factors, cellular responses
CC   to heparin-binding growth factors (e.g. bFGF) and cytokines
CC   (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
CC   cellular susceptibility to certain viral and some bacterial and protozoa
CC   infections, or disintegration of neurodegenerative plaques. The
CC   polynucleotide is also useful in wound healing (e.g. thermal, chemical
CC   or radiation burns), and in the treatment of angiogenesis, restenosis,
CC   atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC   Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC   bacterial or protozoa infections.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 21;  Length 1721;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        160 CGCAAGCACAGGACGTCGTGGACCTG 185


RESULT 11
AAZ39195
ID   AAZ39195 standard; cDNA; 1721 BP.
XX
AC   AAZ39195;
```

```
XX
DT   02-MAR-2000  (first entry)
XX
DE   Human heparanase encoding cDNA.
XX
KW   Human; heparanase; hpa; genetic modification; expression; anticancer;
KW   angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW   anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW   heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW   metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW   neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW   micrometastasis; autoimmune lesion; kidney failure; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1694
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO9957244-A1.
XX
PD   11-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09256.
XX
PR   01-MAY-1998;   98US-0071618.
PR   02-MAR-1999;   99US-0260038.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Ben-Artzi H,  Ayal-Hershkovitz M,  Yacoby-Zeevi O,  Pecker I,  Peleg Y;
PI   Shlomi Y;
XX
DR   WPI; 2000-062144/05.
DR   P-PSDB; AAY57590.
XX
PT   Engineered cells that express recombinant heparanase, useful
PT   therapeutically, e.g. for treating angiogenesis and to screen for
PT   specific inhibitors, potential anticancer agents  -
XX
PS   Claim 2; Page 106-107; 118pp; English.
XX
CC   The present invention describes genetically modified cells (A) containing
CC   a polynucleotide (I) that encodes a polypeptide with heparanase activity,
CC   and express recombinant heparanase (II). Heparanase cleaves heparan
CC   sulphate (HS) at specific intrachain sites, resulting in release of
CC   heparin-binding growth factors, enzymes and proteins that are sequestered
CC   by HS in basement membranes, extracellular matrix or cell surfaces. It
CC   may also be implicated in tumour angiogenesis and metastases. (II) is
CC   potentially useful in wound healing and for treating angiogenesis,
CC   restenosis, atherosclerosis, inflammation, neurodegeneration, viral
CC   infection and cystic fibrosis. It can also be used to neutralise heparin
CC   (an alternative to protamine) and to screen for specific inhibitors
CC   (potentially useful for treating cancer and metastases). Antibodies
CC   raised against (II) are used for immunodetection and diagnosis of
CC   micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
CC   in large quantities, in a form that is homogeneously processed and
CC   activated/neutralised by a dedicated protease. The present sequence
CC   encodes human heparanase.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 21;  Length 1721;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        160 CGCAAGCACAGGACGTCGTGGACCTG 185
```

```
RESULT 12
AAZ33290
ID   AAZ33290 standard; cDNA; 1721 BP.
XX
AC   AAZ33290;
XX
DT   21-FEB-2000  (first entry)
XX
DE   Human heparanase nucleotide sequence.
XX
KW   Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW   antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW   metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW   mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW   inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW   autoimmune disease; anticancer; kidney disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1694
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO9957153-A1.
XX
PD   11-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09255.
XX
PR   01-MAY-1998;   98US-0071739.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Friedman Y,  Perets T;
XX
DR   WPI; 2000-052944/04.
DR   P-PSDB; AAY52990.
XX
PT   Heparanase-specific molecular probes useful for diagnosis and
PT   treatment, e.g. of tumors, and for targeted drug delivery  -
XX
PS   Example; Page 82-84; 90pp; English.
XX
CC   The present invention describes heparanase-specific molecular probes,
CC   useful for methods of detecting heparanase in situ. The probes and
CC   anti-heparanase antibodies are used to detect or quantify the expression
CC   of heparanase, for diagnosis and monitoring of diseases (especially
CC   metastasis), for treatment of heparanase-associated diseases (e.g.
CC   tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC   mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC   metastases) derived from liver, prostate, bladder, breast, ovary,
CC   cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC   disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC   syndrome, sepsis and inflammatory or autoimmune disease), for targeted
CC   drug delivery (e.g. of anticancer agents) and as research reagents.
CC   The present sequence encodes human heparanase, which is used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 21;  Length 1721;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        160 CGCAAGCACAGGACGTCGTGGACCTG 185

RESULT 13
```

```
AAA91112
ID   AAA91112 standard; DNA; 1721 BP.
XX
AC   AAA91112;
XX
DT   20-APR-2001  (first entry)
XX
DE   Human heparanase, coding sequence fragment isolated from EST clone.
XX
KW   Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW   atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW   neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW   gene therapy; mouse; expressed sequence tag; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100643-A2.
XX
PD   04-JAN-2001.
XX
PF   19-JUN-2000; 2000WO-IL00358.
XX
PR   25-JUN-1999;   99US-0140801.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI   Pecker I,  Michal I,  Itzhaki H;
XX
DR   WPI; 2001-137930/14.
XX
PT   New polynucleotides and polypeptides that are distantly homologous to
PT   heparanase, useful in wound healing, as well as in gene therapy
PT   protocols for angiogenesis, restenosis, atherosclerosis, or
PT   inflammation -
XX
PS   Example 1; Page 67; 67pp; English.
XX
CC   This sequence represents a human heparanase coding sequence clone,
CC   isolated from an EST clone. The invention relates to heparanase DNA
CC   and protein sequences. The heparanase DNA and protein sequences are
CC   useful in wound healing, angiogenesis, restenosis, atherosclerosis,
CC   inflammation, pulmonary diseases, neurodegenerative diseases (such as
CC   Scrape, Alzheimer's disease, and Creutzfeldt-Jakob disease) or viral
CC   infections. The heparanase coding sequence is particularly useful in gene
CC   therapy.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 22;  Length 1721;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        160 CGCAAGCACAGGACGTCGTGGACCTG 185


RESULT 14
AAF93788
ID   AAF93788 standard; cDNA; 1722 BP.
XX
AC   AAF93788;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0090.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88361.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 89; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 1722 BP; 449 A; 414 C; 412 G; 447 T; 0 other;

  Query Match             75.4%;  Score 19.6;  DB 22;  Length 1722;
  Best Local Similarity   84.6%;  Pred. No. 42;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        161 CGCAAGCACAGGACGTCGTGGACCTG 186


RESULT 15
AAX37260
ID   AAX37260 standard; DNA; 1723 BP.
XX
AC   AAX37260;
XX
DT   21-JUL-1999  (first entry)
XX
DE   Seq ID No: 14 of WO9921975.
XX
KW   Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW   metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW   arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW   human; HSPG; ss.
XX
OS   Homo sapiens.
```

XX
PN WO9921975-A1.
XX
PD 06-MAY-1999.
XX
PF 28-OCT-1998; 98WO-AU00898.
XX
PR 09-DEC-1997; 97AU-0000812.
PR 28-OCT-1997; 97AU-0000062.
XX
PA (AUSU) UNIV AUSTRALIAN NAT.
XX
PI Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;
XX
DR WPI; 1999-312956/26.
DR P-PSDB; AAY17083.
XX
PT Polynucleotides encoding mammalian endoglucuronidases, especially
PT heparanases, useful to promote wound healing
XX
PS Claim 11; Page 76-79; 112pp; English.
XX
CC The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG.
XX
SQ Sequence 1723 BP; 461 A; 407 C; 412 G; 443 T; 0 other;

Query Match 75.4%; Score 19.6; DB 20; Length 1723;
Best Local Similarity 84.6%; Pred. No. 42;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        149 CGCAAGCACAGGACGTCGTGGACCTG 174
```

Search completed: February 16, 2004, 09:18:10
Job time : 13.7569 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 08:49:50 ; Search time 100.997 Seconds
(without alignments)
6256.802 Million cell updates/sec

Title: US-10-676-079-4
Perfect score: 26
Sequence: 1 cgcatatgcaggacgtcgtggacctg 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19.8	76.2	677	14	CD044218	CD044218 psHB013xH
	2	19.6	75.4	881	14	CB988510	CB988510 AGENCOURT
	3	19.6	75.4	924	13	BQ691142	BQ691142 AGENCOURT
	4	19.6	75.4	1185	9	AL552174	AL552174 AL552174
	5	19.6	75.4	1201	9	AL545270	AL545270 AL545270
	6	19.4	74.6	850	29	CNS0448C	AL273765 Tetraodon
	7	19.4	74.6	871	12	BI754736	BI754736 603025450
	8	19.2	73.8	535	28	BH051472	BH051472 RPCI-24-3
	9	19.2	73.8	613	28	AZ362831	AZ362831 1M0108O02
c	10	19.2	73.8	617	12	BJ008199	BJ008199 BJ008199
c	11	19.2	73.8	627	12	BJ487056	BJ487056 BJ487056
	12	19.2	73.8	711	12	BJ533801	BJ533801 BJ533801
c	13	19	73.1	1012	29	CNS026GG	AL183337 Tetraodon
c	14	18.6	71.5	240	9	AA398146	AA398146 zt88h02.r
	15	18.6	71.5	374	9	AA821887	AA821887 vp23a05.r
c	16	18.6	71.5	434	12	BG960581	BG960581 CM0-CT081
c	17	18.6	71.5	520	9	AL750507	AL750507 AL750507
c	18	18.6	71.5	674	9	AL750664	AL750664 AL750664
	19	18.6	71.5	1177	11	AK009597	AK009597 Mus muscu
c	20	18.6	71.5	1396	14	CD387962	CD387962 AGENCOURT
	21	18.6	71.5	1798	10	BG110194	BG110194 602279842
c	22	18.6	71.5	2666	11	BC030655	BC030655 Homo sapi
c	23	18.4	70.8	196	9	AA243145	AA243145 zs13h05.s
c	24	18.4	70.8	226	13	BY000868	BY000868 BY000868
c	25	18.4	70.8	248	13	BY063809	BY063809 BY063809
c	26	18.4	70.8	303	9	AA811576	AA811576 oa99h05.s
c	27	18.4	70.8	358	9	AA769475	AA769475 nz31b06.s
c	28	18.4	70.8	364	9	AI143664	AI143664 qf21b08.x
c	29	18.4	70.8	367	13	BY072622	BY072622 BY072622
c	30	18.4	70.8	410	9	AA732339	AA732339 nz95h04.s
c	31	18.4	70.8	577	10	AW977606	AW977606 EST389715
c	32	18.4	70.8	707	14	BY735903	BY735903 BY735903
	33	18.4	70.8	730	14	CB031220	CB031220 TgESTzyd4
c	34	18.4	70.8	761	12	BI649991	BI649991 603296650
c	35	18.4	70.8	854	10	BF581403	BF581403 602100804
c	36	18.4	70.8	884	12	BI104061	BI104061 602890362
	37	18.4	70.8	2078	11	BC029952	BC029952 Homo sapi
c	38	18.2	70.0	285	10	BB181339	BB181339 BB181339
	39	18.2	70.0	382	9	AW288247	AW288247 SWOv3MCAM
c	40	18.2	70.0	423	14	CB131946	CB131946 K-EST0182
c	41	18.2	70.0	530	14	CB136264	CB136264 K-EST0188
	42	18.2	70.0	622	28	AZ087080	AZ087080 RPCI-23-2
c	43	18.2	70.0	626	13	BU110645	BU110645 603127747
	44	18.2	70.0	754	10	BE897409	BE897409 601439313
	45	18.2	70.0	759	12	BM018081	BM018081 603645645

ALIGNMENTS

RESULT 1
CD044218
LOCUS CD044218 677 bp mRNA linear EST 09-MAY-2003
DEFINITION psHB013xH08f_181768 psHB: Infected hypocotyl soybean host. 48 hrs post infection Phytophthora sojae cDNA clone sHB013H08 5, mRNA sequence.
ACCESSION CD044218
VERSION CD044218.1 GI:30497811
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
REFERENCE 1 (bases 1 to 677)
AUTHORS Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
TITLE USDA-IFAFS: Expression of Phytophthora sojae genes during infection and propagation
JOURNAL Unpublished
COMMENT Contact: Tyler B
Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtyler@vt.edu
PCR PRimers
FORWARD: BK reverse
Plate: 013 row: H column: 08

```
          Seq primer: BK reverse
          High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sHB013H08"
                     /tissue_type="infected host tissue"
                     /cell_line="P6497"
                     /dev_stage="48 hour post infection"
                     /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
                     post infection"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI;
                     USDA-IFAFS:Expression of Phytophthora sojae genes during
                     infection and propagation."
BASE COUNT      129 a    244 c    198 g    106 t
ORIGIN

  Query Match             76.2%;  Score 19.8;  DB 14;  Length 677;
  Best Local Similarity   91.3%;  Pred. No. 6.3e+02;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 ATATGCAGGACGTCGTGGACCTG 26
              | |||||||||||||||| ||||
Db        339 AGATGCAGGACGTCGTGGGCCTG 361


RESULT 2
CB988510
LOCUS       CB988510                881 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  AGENCOURT_13905817 NIH_MGC_147 Homo sapiens cDNA clone
            IMAGE:30340461 5', mRNA sequence.
ACCESSION   CB988510
VERSION     CB988510.1  GI:30283030
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Stefan Hansson
             cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
            and advice from Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM370  row: f  column: 22
            High quality sequence stop: 664.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30340461"
                     /tissue_type="Human Placenta"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_147"
                     /note="Organ: placenta; Vector: pBluescriptR; Site_1:
                     alI-XhoI; Site_2: BamH; Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: This is
                     a NIH_MGC library."
BASE COUNT      200 a    244 c    229 g    208 t
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 14;  Length 881;
  Best Local Similarity   84.6%;  Pred. No. 8.2e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        219 CGCAAGCACAGGACGTCGTGGACCTG 244


RESULT 3
BQ691142
LOCUS       BQ691142                924 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8343629 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250265
            5', mRNA sequence.
ACCESSION   BQ691142
VERSION     BQ691142.1  GI:21816458
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2393  row: a  column: 18
            High quality sequence stop: 710.
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6250265"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      203 a    271 c    227 g    219 t      4 others
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 13;  Length 924;
  Best Local Similarity   84.6%;  Pred. No. 8.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        232 CGCAAGCACAGGACGTCGTGGACCTG 257


RESULT 4
AL552174
LOCUS       AL552174                1185 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL552174 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
```

clone CS0DI059YN15 5-PRIME, mRNA sequence.
ACCESSION AL552174
VERSION AL552174.2 GI:31273990
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1185)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 15, 2001 this sequence version replaced gi:12890820.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2469.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI059CG08QP1&cluster=2469.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI059CG08QP1.
FEATURES Location/Qualifiers
source 1. .1185
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI059YN15"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 280 a 293 c 293 g 280 t 39 others
ORIGIN

Query Match 75.4%; Score 19.6; DB 9; Length 1185;
Best Local Similarity 84.6%; Pred. No. 8.9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGCATATGCAGGACGTCGTGGACCTG 26
|||| ||||||||||||||||||
Db 226 CGCAAGCACAGGACGTCGTGGACCTG 251

RESULT 5
AL545270
LOCUS AL545270 1201 bp mRNA linear EST 31-MAY-2003
DEFINITION AL545270 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI028YF04 5-PRIME, mRNA sequence.
ACCESSION AL545270
VERSION AL545270.2 GI:31267106
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 15, 2001 this sequence version replaced gi:12877751.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2469.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI028DC02QP1&cluster=2469.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI028DC02QP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI028YF04"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 292 a 282 c 305 g 279 t 43 others
ORIGIN

Query Match 75.4%; Score 19.6; DB 9; Length 1201;
Best Local Similarity 84.6%; Pred. No. 8.9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGCATATGCAGGACGTCGTGGACCTG 26
|||| ||||||||||||||||||
Db 219 CGCAAGCACAGGACGTCGTGGACCTG 244

RESULT 6
CNS0448C
LOCUS CNS0448C 850 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 081G03 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL273765
VERSION AL273765.1 GI:7996028
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 850)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers

```
     source          1. .850
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="081G03"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG081AD02SP1~end :
                     PUC-Ori"
BASE COUNT      156 a    251 c    258 g    179 t      6 others
ORIGIN

  Query Match             74.6%;  Score 19.4;  DB 29;  Length 850;
  Best Local Similarity   95.2%;  Pred. No. 9.7e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 ATGCAGGACGTCGTGGACCTG 26
              ||||||||| ||||||||||
Db         54 ATGCAGGACATCGTGGACCTG 74


RESULT 7
BI754736
LOCUS       BI754736                 871 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603025450F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5195845 5',
            mRNA sequence.
ACCESSION   BI754736
VERSION     BI754736.1  GI:15746314
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 871)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11490  row: c  column: 14
            High quality sequence stop: 644.
FEATURES             Location/Qualifiers
     source          1. .871
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5195845"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      174 a    268 c    277 g    152 t
ORIGIN

  Query Match             74.6%;  Score 19.4;  DB 12;  Length 871;
  Best Local Similarity   95.2%;  Pred. No. 9.8e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 ATGCAGGACGTCGTGGACCTG 26
              ||||||||| ||||||||||
Db        558 ATGCAGGACTTCGTGGACCTG 578


RESULT 8
BH051472
LOCUS       BH051472                 535 bp    DNA     linear   GSS 17-JUL-2001
DEFINITION  RPCI-24-307N24.TJ RPCI-24 Mus musculus genomic clone RPCI-24-307N24
            , genomic survey sequence.
ACCESSION   BH051472
VERSION     BH051472.1  GI:14843016
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-24-307N24.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 307  row: N  column: 24
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-307N24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      114 a    148 c    166 g    107 t
ORIGIN

  Query Match             73.8%;  Score 19.2;  DB 28;  Length 535;
  Best Local Similarity   87.5%;  Pred. No. 1e+03;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 CATATGCAGGACGTCGTGGACCTG 26
              ||| |||||||| | |||||||||
Db        150 CATGTGCAGGACATGGTGGACCTG 173


RESULT 9
AZ362831
LOCUS       AZ362831                 613 bp    DNA     linear   GSS 02-OCT-2000
DEFINITION  1M0108O02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0108O02 F, genomic survey sequence.
ACCESSION   AZ362831
VERSION     AZ362831.1  GI:10476531
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0108  row: O  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0108O02"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      153 a    162 c    187 g    111 t
ORIGIN

  Query Match             73.8%;  Score 19.2;  DB 28;  Length 613;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 CATATGCAGGACGTCGTGGACCTG 26
              ||| |||||||| | |||||||||
Db        438 CATGTGCAGGACATGGTGGACCTG 461


RESULT 10
BJ008199/c
LOCUS       BJ008199                617 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ008199 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA116G09 5',
            mRNA sequence.
ACCESSION   BJ008199
VERSION     BJ008199.1  GI:17366522
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA116G09"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
BASE COUNT      179 a    154 c    202 g     82 t
ORIGIN

  Query Match             73.8%;  Score 19.2;  DB 12;  Length 617;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GCATATGCAGGACGTCGTGGACCT 25
              || | | ||||||||||||||||
Db        201 GCCTGTTCAGGACGTCGTGGACCT 178


RESULT 11
BJ487056/c
LOCUS       BJ487056                627 bp    mRNA    linear   EST 09-AUG-2002
DEFINITION  BJ487056 MF01FSA cDNA Oryzias latipes cDNA clone MF01FSA002C15 5',
            mRNA sequence.
ACCESSION   BJ487056
VERSION     BJ487056.1  GI:22165805
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="d-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01FSA002C15"
                     /sex="mixture of female and male"
```

```
                     /tissue_type="whole embryo"
                     /dev_stage="fry stage 40"
                     /clone_lib="MF01FSA cDNA"
BASE COUNT      177 a    168 c    189 g     92 t      1 others
ORIGIN

  Query Match             73.8%;  Score 19.2;  DB 12;  Length 627;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GCATATGCAGGACGTCGTGGACCT 25
              || | | ||||||||||||||||||
Db        314 GCCTGTTCAGGACGTCGTGGACCT 291


RESULT 12
BJ533801
LOCUS       BJ533801                711 bp    mRNA    linear   EST 09-AUG-2002
DEFINITION  BJ533801 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB018B06 3',
            mRNA sequence.
ACCESSION   BJ533801
VERSION     BJ533801.1  GI:22192613
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSB018B06"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSB cDNA"
BASE COUNT      119 a    194 c    208 g    188 t      2 others
ORIGIN

  Query Match             73.8%;  Score 19.2;  DB 12;  Length 711;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GCATATGCAGGACGTCGTGGACCT 25
              || | | ||||||||||||||||||
Db        322 GCCTGTTCAGGACGTCGTGGACCT 345


RESULT 13
CNS026GG/c
LOCUS       CNS026GG                1012 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            240N06 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL183337
VERSION     AL183337.1  GI:7821441
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1012)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="240N06"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG240DG03LP1-end : T7"
BASE COUNT      174 a    323 c    300 g    206 t      9 others
ORIGIN

  Query Match             73.1%;  Score 19;  DB 29;  Length 1012;
  Best Local Similarity   90.5%;  Pred. No. 1.5e+03;
  Matches   19;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 ATGCAGGACGTCGTGGACCTG 26
              ||||||||| ||||||||||:
Db        306 ATGCAGGACATCGTGGACCTS 286


RESULT 14
AA398146/c
LOCUS       AA398146                240 bp    mRNA    linear   EST 12-AUG-1997
DEFINITION  zt88h02.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:729459
            5', mRNA sequence.
ACCESSION   AA398146
VERSION     AA398146.1  GI:2051273
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
```

```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1694   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 130.
FEATURES              Location/Qualifiers
     source           1. .240
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:5925988"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:729459"
                      /sex="male"
                      /lab_host="DH10B"
                      /clone_lib="Soares_testis_NHT"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from Clontech Laboratories
                      , Inc., and primed with  a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization to Cot5, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       51 a     72 c     50 g     67 t
ORIGIN

  Query Match             71.5%;  Score 18.6;  DB 9;  Length 240;
  Best Local Similarity   84.0%;  Pred. No. 1.4e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 GCATATGCAGGACGTCGTGGACCTG 26
              ||| ||| || |||| |||||||||
Db        227 GCAAATGGAGCACGTGGTGGACCTG 203


RESULT 15
AA821887
LOCUS       AA821887                374 bp    mRNA    linear   EST 17-FEB-1998
DEFINITION  vp23a05.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
            clone IMAGE:1077488 5', mRNA sequence.
ACCESSION   AA821887
VERSION     AA821887.1  GI:2891755
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:591784
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 371.
FEATURES              Location/Qualifiers
     source           1. .374
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1077488"
                      /tissue_type="diaphragm"
                      /dev_stage="adult"
                      /lab_host="SOLR (kanamycin resistant)"
                      /clone_lib="Stratagene mouse diaphragm (#937303)"
                      /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
                      prepared from diaphragm muscle.  Primer: Oligo dT. Average
                      insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       72 a     113 c     130 g     59 t
ORIGIN

  Query Match             71.5%;  Score 18.6;  DB 9;  Length 374;
  Best Local Similarity   84.0%;  Pred. No. 1.6e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 GCATATGCAGGACGTCGTGGACCTG 26
              || | ||||||| ||| ||||||||
Db         99 GCCTGTGCAGGAGGTCCTGGACCTG 123


Search completed: February 16, 2004, 13:40:54
Job time : 105.997 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:56:25 ; Search time 124.136 Seconds
(without alignments)
8568.399 Million cell updates/sec

Title: US-10-676-079-4
Perfect score: 26
Sequence: 1 cgcatatgcaggacgtcgtggacctg 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	100.0	26	6	AR125605	AR125605 Sequence
2	26	100.0	26	6	AR194191	AR194191 Sequence
3	26	100.0	26	6	AR221287	AR221287 Sequence
4	26	100.0	26	6	AR243205	AR243205 Sequence
5	26	100.0	26	6	AR287437	AR287437 Sequence
6	20	76.9	30	6	AX557108	AX557108 Sequence
c 7	19.8	76.2	10328	1	AE006005	AE006005 Caulobact
8	19.6	75.4	474	6	AX136496	AX136496 Sequence
9	19.6	75.4	474	6	BD123736	BD123736 Secretory
10	19.6	75.4	605	6	AX557103	AX557103 Sequence
11	19.6	75.4	1419	6	AX557106	AX557106 Sequence
12	19.6	75.4	1593	6	AR210040	AR210040 Sequence
13	19.6	75.4	1593	6	BD136761	BD136761 Human pla
14	19.6	75.4	1694	9	AF152376	AF152376 Homo sapi
15	19.6	75.4	1713	6	AR156691	AR156691 Sequence
16	19.6	75.4	1713	6	AX034643	AX034643 Sequence
17	19.6	75.4	1721	6	AR080679	AR080679 Sequence
18	19.6	75.4	1721	6	AR080680	AR080680 Sequence
19	19.6	75.4	1721	6	AR125603	AR125603 Sequence
20	19.6	75.4	1721	6	AR125604	AR125604 Sequence
21	19.6	75.4	1721	6	AR194189	AR194189 Sequence
22	19.6	75.4	1721	6	AR194190	AR194190 Sequence
23	19.6	75.4	1721	6	AR221285	AR221285 Sequence
24	19.6	75.4	1721	6	AR221286	AR221286 Sequence
25	19.6	75.4	1721	6	AR243203	AR243203 Sequence
26	19.6	75.4	1721	6	AR243204	AR243204 Sequence
27	19.6	75.4	1721	6	AR287435	AR287435 Sequence
28	19.6	75.4	1721	6	AR287436	AR287436 Sequence
29	19.6	75.4	1721	6	BD074427	BD074427 Polynucle
30	19.6	75.4	1721	6	BD074428	BD074428 Polynucle
31	19.6	75.4	1722	6	AX136167	AX136167 Sequence
32	19.6	75.4	1722	6	BD123536	BD123536 Secretory
33	19.6	75.4	1722	9	AK075400	AK075400 Homo sapi
34	19.6	75.4	1723	6	AR156692	AR156692 Sequence
35	19.6	75.4	1723	6	AX034645	AX034645 Sequence
36	19.6	75.4	1724	6	AX147946	AX147946 Sequence
37	19.6	75.4	1724	9	AF165154	AF165154 Homo sapi
38	19.6	75.4	1758	9	AF144325	AF144325 Homo sapi
39	19.6	75.4	1810	9	BC051321	BC051321 Homo sapi
40	19.6	75.4	1899	6	BD074430	BD074430 Polynucle
41	19.6	75.4	1899	6	BD074431	BD074431 Polynucle
42	19.6	75.4	3726	6	AR235866	AR235866 Sequence
43	19.6	75.4	3726	6	AX019348	AX019348 Sequence
44	19.6	75.4	3726	6	BD131218	BD131218 Human hep
45	19.6	75.4	3726	9	AF155510	AF155510 Homo sapi

ALIGNMENTS

RESULT 1
AR125605
LOCUS AR125605 26 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 4 from patent US 6177545.
ACCESSION AR125605
VERSION AR125605.1 GI:14111667
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 26)
AUTHORS Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
TITLE Heparanase specific molecular probes and their use in research and medical applications
JOURNAL Patent: US 6177545-A 4 23-JAN-2001;

```
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="unknown"
BASE COUNT        5 a      7 c      9 g      5 t
ORIGIN

  Query Match             100.0%;  Score 26;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.3;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||||||||||||||||||||||||
Db          1 CGCATATGCAGGACGTCGTGGACCTG 26


RESULT 2
AR194191
LOCUS       AR194191                  26 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 4 from patent US 6348344.
ACCESSION   AR194191
VERSION     AR194191.1  GI:20240783
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M.,
            Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Schlomi,Y.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6348344-A 4 19-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="unknown"
BASE COUNT        5 a      7 c      9 g      5 t
ORIGIN

  Query Match             100.0%;  Score 26;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.3;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||||||||||||||||||||||||
Db          1 CGCATATGCAGGACGTCGTGGACCTG 26


RESULT 3
AR221287
LOCUS       AR221287                  26 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 4 from patent US 6426209.
ACCESSION   AR221287
VERSION     AR221287.1  GI:23328258
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Ayal-Hershkovitz,M., Pecker,I. and Yacoby-Zeevi,O.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6426209-A 4 30-JUL-2002;
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="unknown"
BASE COUNT        5 a      7 c      9 g      5 t
ORIGIN

  Query Match             100.0%;  Score 26;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.3;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||||||||||||||||||||||||
Db          1 CGCATATGCAGGACGTCGTGGACCTG 26


RESULT 4
AR243205
LOCUS       AR243205                  26 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 4 from patent US 6475763.
ACCESSION   AR243205
VERSION     AR243205.1  GI:27290320
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M.,
            Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Shlomi,Y.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6475763-A 4 05-NOV-2002;
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="unknown"
BASE COUNT        5 a      7 c      9 g      5 t
ORIGIN

  Query Match             100.0%;  Score 26;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.3;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||||||||||||||||||||||||
Db          1 CGCATATGCAGGACGTCGTGGACCTG 26


RESULT 5
AR287437
LOCUS       AR287437                  26 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 4 from patent US 6531129.
ACCESSION   AR287437
VERSION     AR287437.1  GI:29725131
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6531129-A 4 11-MAR-2003;
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="unknown"
BASE COUNT        5 a      7 c      9 g      5 t
ORIGIN

  Query Match             100.0%;  Score 26;  DB 6;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.3;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||||||||||||||||||||||||
Db          1 CGCATATGCAGGACGTCGTGGACCTG 26


RESULT 6
AX557108
LOCUS       AX557108                  30 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 6 from Patent WO0244353.
ACCESSION   AX557108
VERSION     AX557108.1  GI:25900161
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

```
          artificial sequences.
REFERENCE   1
  AUTHORS   Wolffe,A.P.
  TITLE     Human heparanase gene regulatory sequences
  JOURNAL   Patent: WO 0244353-A 6 06-JUN-2002;
          Sangamo Biosciences Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="primer HP-6"
BASE COUNT        4 a      9 c      9 g      8 t
ORIGIN

  Query Match             76.9%;  Score 20;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 TGCAGGACGTCGTGGACCTG 26
              ||||||||||||||||||||
Db          2 TGCAGGACGTCGTGGACCTG 21


RESULT 7
AE006005/c
LOCUS       AE006005               10328 bp    DNA     linear   BCT 12-JUN-2002
DEFINITION  Caulobacter crescentus CB15 section 331 of 359 of the complete
          genome.
ACCESSION   AE006005 AE005673
VERSION     AE006005.1  GI:13425180
KEYWORDS    .
SOURCE      Caulobacter crescentus CB15
  ORGANISM  Caulobacter crescentus CB15
          Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
          Caulobacteraceae; Caulobacter.
REFERENCE   1  (bases 1 to 10328)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
          Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R.,
          Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
          Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
          Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
          Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
          Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
          Fraser,C.M.
  TITLE     Complete genome sequence of Caulobacter crescentus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
  MEDLINE   21173698
   PUBMED   11259647
REFERENCE   2  (bases 1 to 10328)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
          Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
          Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
          Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
          Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
          Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
          Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
          Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10328
                     /organism="Caulobacter crescentus CB15"
                     /mol_type="genomic DNA"
                     /strain="CB15"
                     /db_xref="taxon:190650"
     gene            complement(90. .2633)
                     /gene="CC3461"
     CDS             complement(90. .2633)
                     /gene="CC3461"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="TonB-dependent receptor"
                     /protein_id="AAK25423.1"
                     /db_xref="GI:13425181"
                     /translation="MITVLSVSNALILMPSFRVDFHQDEGEPCRCVGSGNSKNAKNTR
                     EEIMTIVSSRGRARALFGASALSSLLGAGLVALPSVATAQQADPNAIAEIIVTATKRD
                     ATIQDIPFSINAQTEKDIQRSGAVTLEDLSRNVAGLTIQNLGPGQSQVSVRGVSAGQV
                     VRDQPGVKEQVGVYLDESVISLSLFTPDVDLFDLNRVETLRGPQGTLFGSGSVGGTIR
                     YITNQPKLGVSEGTFEANANLVDGDAFGGHVKGAVNIPISDKIAMRAVGYLTQYGGFI
                     DARKEGGKVDKNVNDGTRRGGRLSFYIEPNEQVNFTPRVVYQEIRAGGFNRQETFNLF
                     ANNNTTTRPKIQLGEREQYLLLDERFSDNTLLADMTANLKFDAATLTSVTSFISRDIT
                     VSRDASALTGSVSVDLGYPAAAVLLPSNLRDTTDLEQFTQELRLASTTDGPLQWVIGG
                     FYSKVDRVYNQRLPTTGYDTYTDAVLGAGTSAAVANGFGKDSPYNAYLPYDIKQKALF
                     GELNYTVGKLTATAGGRYYDFSETRQFKSGGLFANGDNRTDKTSSDGFTPRFLLSYEA
                     SDTVTFNAQASKGFRLGGVNDPLNIPLCSAQDRAIFGGYQNYDDETLWNYEGGVKSRF
                     GRVTLNAAAFYTDIKNLQTTLDAGSCSSRVVFNVPKAHTKGVEAELTARLANGLDIGL
                     SGSLLEAEFDSTVKDGTGAVIGGIREGNRLPSVPKFQVSFDVTYSKEVRPGVNGYLKA
                     SLQHVGNRFTQASDQENNPRLFVSNLAFGGATGRVPTAVNLALPSYEIVNLSAGLEMQ
                     DGVDLTLYVNNLFDENALLSFDRERGGRARLGYAVGQPRTMGVTVRKSF"
     gene            2758. .3945
                     /gene="CC3462"
     CDS             2758. .3945
                     /gene="CC3462"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="thiolase family protein"
                     /protein_id="AAK25424.1"
                     /db_xref="GI:13425182"
                     /translation="MREAVIVSYARTGLAKSVRGGFNNTHGAAMAGHAIQHAVSRAGL
                     EGAEVEDVVLGCGGPEGATGMNVARNAAMWAGLPVTTSGQTINRFCSSGLQAIATAAN
                     YVRNDGANVAIGGGVESISLVNAGGHMNRFHITEEKLMQTHPALWMAMIDTADIVAKR
                     YNVSREYQDEYALRSQQRIAAAQAAGLFKDEIVPMATKMKVVNKETKEESFVDYVVDK
                     DECNRADTTLEGLASLKPVMGEGKFITAGNASQLSDGAAAVVVMEAKEAEKRGLTPLG
                     AFRGFAVAGCEPDEMGIGPVFAVPRLLERHGLKVDDIDIWELNEAFASQCLYSRDRLG
                     IDPEKYNVNGGSIAIGHPFGMTGARCAGHLLLEGKRRKAKLGVVTMCIGGGMGAAGLF
                     EIF"
     gene            complement(3998. .4279)
                     /gene="CC3463"
     CDS             complement(3998. .4279)
                     /gene="CC3463"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK25425.1"
                     /db_xref="GI:13425183"
                     /translation="MWELTRMEILVLRQAQDEDFLRTEANSRPHPELVEGRGRLLRGD
                     QHPTSKQPKERPETAKAVSGLFQSSTLSDGCAVRNQKISNRPAAPLWIA"
     gene            complement(4343. .7246)
                     /gene="CC3464"
     CDS             complement(4343. .7246)
                     /gene="CC3464"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA polymerase I"
                     /protein_id="AAK25426.1"
                     /db_xref="GI:13425184"
                     /translation="MLPAMTDAAAPPLPNASELTQDGPAVRLFLVDGSAYLFRAYHAL
                     PPLTRKSDGLPVGAVQGFCNMLWKLLRDMQGDTPTHLAVIWDHSEKTFRNALYDQYKA
                     HRPPPPEDLIPQFPLVREATLAFGVPAIELPGYEADDLIAAYACKARDIGGEAIIVSS
                     DKDLMQLVGDGVSMYDPMKGVRIEREQVFEKFGVYPEKVVDVQALCGDSVDNVPGAPG
                     IGIKTAAQLITEYGDLDTLLARAGEIKQPKRRETLINFADQIRLSRALVKLDCDTPLP
                     QPLDALTVREPDKEALAAFLEQMEFRSLARRVGDGSAAATPGTLDRPAAPPKAPVVSV
                     SYMGAAARAAAHPVEPVKIDHAAYACVRDLATLKAWVAKATDKGLVAFDTETDALSSA
                     TAGLCGVSLAIAPGEACYIPISHCEKADGLAFEAPADIEQIPLADVIATLKPLLEDPA
                     VLKVAQNAKYDIAVLARHGIQVAPIEDTMLISYVLEAGLHGHGMDELSELHLGHKPIP
                     FKQVAGSGKGQISFKHVALPEATAYAAEDADVTLRLYHHLKPQLARASLSTVYETLER
                     PMPAVLAMMENNGVRVDPEALRLLSNEFSLRMAQFEARAQELVGRPFNLGSPKQIGDV
                     LFGEMQMKGGKKTATGQWSTDSDVLESLALEHELPRVLLDWRQLSKLKGTYTENLIAA
                     IAPSTGRVHTSYALAATTTGRLSSSDPNLQNIPVRTEEGRKIRKAFVAAPGHVLISAD
                     YSQIELRLLAHIGDIPQLKKAFQEGLDIHAMTASEMFDTPIEGMDPMIRRRAKAINFG
```

```
                IVYGISAFGLANQLGISQGEAGAYIKTYFERFPGIQAYMDATKAFVREHGYVTTIFGR
                KINIPDIEAKSAAHRQFAERAAINAPIQGAAADVMRRAMIRMPVALEAAGLSTRMLLQ
                VHDELVFEAPEAEAERACAVVRAVMEKAAEPAVALSIPLTVEARAASNWDEAH"
     gene       7355. .9202
                /gene="CC3465"
     CDS        7355. .9202
                /gene="CC3465"
                /note="identified by match to  protein family HMM"
                /codon_start=1
                /transl_table=11
                /product="ATP-dependent DNA helicase RecQ"
                /protein_id="AAK25427.1"
                /db_xref="GI:13425185"
                /translation="MYVPPESPELDRARDVLRRTFGHADFRGLQAGVIHELLTGHSAM
                AVLPTGGGKSLCYQIPSLIRPGLGLVISPLIALMADQVQGLRQAGVAAERLDSNISMD
                ERSDIWRRIDAGEVDLLYLSPEGLMQPWMLDRLSRTPLALIAVDEAHCVSQWGHDFRP
                EYRMLGRLAELFPDAPRLAVTATADARTRDDIRAELRLQGAAEFVDSFARPELALSAE
                RKRGKGHDRVVELVLERPGRSGVIYAGSRDGTEKLAERLNAEGVPALAYHAGLDKAVR
                ARRLEDFLEADAAVMVATIAFGMGVDKPDVRYVIHADPPAAIEAYWQEVGRAGRDGQP
                AEGITLYGSADMAWAARRIETREAPDEVKQVQSRKLRQFYAMLEGVTCRAAAVRRYFG
                EEGVGHCGVCDICVSPPTGIDATQAAQKALSAVHRLGGRLGRGRVIEHLMGKTKDVTP
                QEAQLPTFGIGREFSQPTWRDLFDTLIFEGLLREDPNDGRPLIGLGDVEGVRQVYRNE
                RKVALRQTTDAPDSGGRAGGGMRKRREGRALTIPAENQLLFEALRSWRKEQAQLQHVP
                PYVIFHDATLAEIAAARPATLAALGKAGGVGQGKLDRYGEAVLKVVREN"
     gene       9247. .9555
                /gene="CC3466"
     CDS        9247. .9555
                /gene="CC3466"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK25428.1"
                /db_xref="GI:13425186"
                /translation="MTHVNSLTPSDITSAVDDIVENTKRAKDAALDAVKSARDTATEV
                AQQTRTFAAEQARAGAAWTTKQAKEAHHVAEDHPMGLALGAFALGLGLGFLISRNLAR
                "
     gene       complement(9792. .10250)
                /gene="CC3467"
     CDS        complement(9792. .10250)
                /gene="CC3467"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAK25429.1"
                /db_xref="GI:13425187"
                /translation="MDVIIELAPSRPETTLAVTRGAARLLVDLGYAPLAEVTLPNGRR
                ADLMALGPKGDVLIVEVKSGLEDFRVDRKWGDYAPFCDAFYFAVAPSFPDGILPEDPG
                LLVADGFGGAVVREAPLTPLAPARRKALTLAFGRLAAWRAAGVNAERLSL"
BASE COUNT     1665 a   3459 c   3503 g   1701 t
ORIGIN

  Query Match           76.2%;  Score 19.8;  DB 1;  Length 10328;
  Best Local Similarity   91.3%;  Pred. No. 5.7e+02;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 ATATGCAGGACGTCGTGGACCTG 26
              | ||||||||||| ||||||||||
Db        247 AGATGCAGGACGGCGTGGACCTG 225


RESULT 8
AX136496
LOCUS       AX136496                 474 bp    DNA     linear   PAT 30-MAY-2001
DEFINITION  Sequence 418 from Patent EP1067182.
ACCESSION   AX136496
VERSION     AX136496.1  GI:14272900
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
            Hayashi,K.
  TITLE     Secretory protein or membrane protein
  JOURNAL   Patent: EP 1067182-A 418 10-JAN-2001;
            Helix Research Institute (JP)
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT       86 a    154 c    127 g    101 t      6 others
ORIGIN

  Query Match           75.4%;  Score 19.6;  DB 6;  Length 474;
  Best Local Similarity   84.6%;  Pred. No. 7.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||||
Db        161 CGCAAGCACAGGACGTCGTGGACCTG 186


RESULT 9
BD123736
LOCUS       BD123736                 474 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Secretory protein or membrane protein.
ACCESSION   BD123736
VERSION     BD123736.1  GI:23218681
KEYWORDS    JP 2002017376-A/245.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
            Hayashi,K.
  TITLE     Secretory protein or membrane protein
  JOURNAL   Patent: JP 2002017376-A 245 22-JAN-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002017376-A/245
            PD   22-JAN-2002
            PF   07-JUL-2000 JP 2000253173
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,YURI KAWAI,TOMOYASU
            PI   SUGIYAMA,
            PI   KOJI HAYASHI
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,C12N15/00,C12N5/00 CC
            Secretory protein or membrane protein
            FH   Key             Location/Qualifiers
            FT   source          1. .474
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT       86 a    154 c    127 g    101 t      6 others
ORIGIN

  Query Match           75.4%;  Score 19.6;  DB 6;  Length 474;
  Best Local Similarity   84.6%;  Pred. No. 7.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||||
Db        161 CGCAAGCACAGGACGTCGTGGACCTG 186


RESULT 10
```

```
AX557103
LOCUS       AX557103                605 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 1 from Patent WO0244353.
ACCESSION   AX557103
VERSION     AX557103.1  GI:25900156
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Wolffe,A.P.
  TITLE     Human heparanase gene regulatory sequences
  JOURNAL   Patent: WO 0244353-A 1 06-JUN-2002;
            Sangamo Biosciences Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="first and second exons of human heparanase gene"
BASE COUNT       96 a    187 c    230 g     92 t
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 6;  Length 605;
  Best Local Similarity   84.6%;  Pred. No. 7.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        468 CGCAAGCACAGGACGTCGTGGACCTG 493


RESULT 11
AX557106
LOCUS       AX557106                1419 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 4 from Patent WO0244353.
ACCESSION   AX557106
VERSION     AX557106.1  GI:25900159
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Wolffe,A.P.
  TITLE     Human heparanase gene regulatory sequences
  JOURNAL   Patent: WO 0244353-A 4 06-JUN-2002;
            Sangamo Biosciences Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1419
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="human heparanase gene in the vicinity of the
                     upstream region"
BASE COUNT      329 a    352 c    452 g    286 t
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 6;  Length 1419;
  Best Local Similarity   84.6%;  Pred. No. 7.2e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db       1282 CGCAAGCACAGGACGTCGTGGACCTG 1307


RESULT 12
AR210040
LOCUS       AR210040                1593 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6387643.
ACCESSION   AR210040
VERSION     AR210040.1  GI:21512167
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Heinrikson,R.Leroy., Fairbanks,M.B. and Mildner,A.M.
  TITLE     Human platelet heparanase polypeptides, polynucleotide molecules
            that encode them, and methods for the identification of compounds
            that alter heparanase activity
  JOURNAL   Patent: US 6387643-A 1 14-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="unknown"
BASE COUNT      426 a    370 c    369 g    428 t
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 6;  Length 1593;
  Best Local Similarity   84.6%;  Pred. No. 7.2e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db         59 CGCAAGCACAGGACGTCGTGGACCTG 84


RESULT 13
BD136761
LOCUS       BD136761                1593 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Human platelet heparanase polypeptide, polynucleotide molecule
            encoding the same and method of identifying compound changing
            heparanase activity.
ACCESSION   BD136761
VERSION     BD136761.1  GI:23231706
KEYWORDS    JP 2002504376-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1593)
  AUTHORS   Heinrikson,R.L., Fairbanks,M.B. and Mildner,A.M.
  TITLE     Human platelet heparanase polypeptide, polynucleotide molecule
            encoding the same and method of identifying compound changing
  JOURNAL   Patent: JP 2002504376-A 1 12-FEB-2002;
            PHARMACIA & UPJOHN CO
COMMENT     OS   Unidentified
            PN   JP 2002504376-A/1
            PD   12-FEB-2002
            PF   18-FEB-1999 JP 2000533569
            PR   24-FEB-1998 US    60/075706,26-MAR-1998 US    60/079401 PI
            ROBERT L HEINRIKSON,MICHAEL B FAIRBANKS,ANA M MILDNER PC
            C12N15/09,C07K16/40,C12N1/21,C12N5/10,C12N9/24,C12Q1/34,C12N15/ PC
             00,C12N5/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            CC   Human platelet heparanase polypeptide, polynucleotide molecule

            CC           encoding
            CC   the same and method of identifying compound changing  CC
                 heparanase activity
            FH   Key             Location/Qualifiers
            FT   source          1. .1593
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1593
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      426 a    370 c    369 g    428 t
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 6;  Length 1593;
  Best Local Similarity   84.6%;  Pred. No. 7.2e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
```

```
              ||||    ||||||||||||||||||
Db         59 CGCAAGCACAGGACGTCGTGGACCTG 84


RESULT 14
AF152376
LOCUS       AF152376               1694 bp    mRNA    linear   PRI 28-JUL-1999
DEFINITION  Homo sapiens heparanase mRNA, complete cds.
ACCESSION   AF152376
VERSION     AF152376.1  GI:5616196
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1694)
  AUTHORS   Kussie,P.H., Hulmes,J.D., Ludwig,D.L., Patel,S., Navarro,E.C.,
            Seddon,A.P., Giorgio,N.A. and Bohlen,P.
  TITLE     Cloning and functional expression of a human heparanase gene
  JOURNAL   Biochem. Biophys. Res. Commun. 261 (1), 183-187 (1999)
  MEDLINE   99335379
   PUBMED   10405343
REFERENCE   2  (bases 1 to 1694)
  AUTHORS   Kussie,P.H., Hulmes,J.D., Ludwig,D., Patel,S., Navarro,E.C.,
            Seddon,A.P., Giorgio,N.A. and Bohlen,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-1999) Protein Chemistry, ImClone Systems Inc.,
            180 Varick Street, New York, NY 10014, USA
FEATURES             Location/Qualifiers
     source          1. .1694
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
     CDS             15. .1646
                     /codon_start=1
                     /product="heparanase"
                     /protein_id="AAD45669.1"
                     /db_xref="GI:5616197"
                     /translation="MLLRSKPALPPPLMLLLLGPLGPLSPGALPRPAQAQDVVDLDFF
                     TQEPLHLVSPSFLSVTIDANLATDPRFLILLGSPKLRTLARGLSPAYLRFGGTKTDFL
                     IFDPKKESTFEERSYWQSQVNQDICKYGSIPPDVEEKLRLEWPYQEQLLLREHYQKKF
                     KNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNI
                     SWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQPRRKTA
                     KMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTR
                     PGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYH
                     LVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTNTDNPRYKEGDL
                     TLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLNGLTLKMVDDQTL
                     PPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI"
BASE COUNT      465 a    398 c    391 g    440 t
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 9;  Length 1694;
  Best Local Similarity   84.6%;  Pred. No. 7.1e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        112 CGCAAGCACAGGACGTCGTGGACCTG 137


RESULT 15
AR156691
LOCUS       AR156691               1713 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 12 from patent US 6242238.
ACCESSION   AR156691
VERSION     AR156691.1  GI:15125395
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1713)
  AUTHORS   Freeman,C.Geoffrey., Hulett,M.Darren., Parish,C.Richard. and
            Hamdorf,B.James.
  TITLE     Isolated nucleic acid molecule encoding mammalian endoglucuronidase
            and uses therefor
  JOURNAL   Patent: US 6242238-A 12 05-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .1713
                     /organism="unknown"
BASE COUNT      460 a    404 c    406 g    443 t
ORIGIN

  Query Match             75.4%;  Score 19.6;  DB 6;  Length 1713;
  Best Local Similarity   84.6%;  Pred. No. 7.1e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CGCATATGCAGGACGTCGTGGACCTG 26
              ||||    ||||||||||||||||||
Db        143 CGCAAGCACAGGACGTCGTGGACCTG 168


Search completed: February 16, 2004, 11:42:59
Job time : 125.136 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:55:00 ; Search time 844.407 Seconds
(without alignments)
5501.769 Million cell updates/sec

Title: US-10-676-079-1
Perfect score: 1721
Sequence: 1 ctagagctttcgactctccg.....atatactagtcctgacactg 1721

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1721	100.0	1721	20	AAX35648	cDNA encoding a hu
	2	1721	100.0	1721	21	AAA75051	cDNA encoding a hu
	3	1721	100.0	1721	21	AAZ39195	Human heparanase e
	4	1721	100.0	1721	21	AAZ33290	Human heparanase n
	5	1721	100.0	1721	22	AAA91112	Human heparanase,
	6	1719.4	99.9	1899	20	AAX35650	cDNA encoding a hu
	7	1719.4	99.9	1899	21	AAA75053	cDNA encoding a hu
	8	1713	99.5	1722	22	AAF93788	Human cDNA encodin
	9	1694.6	98.5	1713	20	AAX37259	Human heparanase e
	10	1688.8	98.1	1723	20	AAX37260	Seq ID No: 14 of W
	11	1686.8	98.0	3726	20	AAX86671	cDNA encoding a hu
	12	1682.6	97.8	1724	22	AAH20940	Human heparanase i
	13	1631.4	94.8	1669	25	ABZ22816	Human heparanase e
	14	1585	92.1	1593	20	AAZ11236	Human pre-prohepar
	15	1535	89.2	1584	24	ABL40753	Chicken signal pep
	16	1092	63.5	2396	21	AAA75081	cDNA encoding a mu
	17	1092	63.5	2396	22	AAA91113	Mouse heparanase c
	18	802.2	46.6	1380	20	AAX37261	Murine heparanase
	19	786.2	45.7	1192	20	AAX37262	Rat heparanase enz
	20	595	34.6	1605	24	ABL40748	Chicken heparanase
c	21	486.8	28.3	553	24	ABL68772	Kidney cancer rela
	22	453	26.3	824	20	AAX35649	3' untranslated re
	23	453	26.3	824	21	AAA75052	Murine EST which i
	24	435.2	25.3	474	22	AAF93984	Primer specific fo
c	25	433.6	25.2	504	22	AAF94131	Primer specific fo
c	26	313.8	18.2	428	25	ABX51804	Bovine EST associa
	27	287.8	16.7	605	24	ABN86000	Human heparanase g
	28	287.8	16.7	1419	24	ABN86003	Human heparanase u
	29	287.8	16.7	44848	21	AAA75080	Nucleotide sequenc
	30	285.2	16.6	2060	22	AAA91097	Human heparanase,
	31	284.8	16.5	1847	24	AAD29202	Human heparanase-2
	32	284.8	16.5	2636	22	AAH22671	Heparanase-like pr
	33	284.6	16.5	2496	24	AAI70849	Human heparanase I
	34	282	16.4	1779	22	AAS13848	DNA encoding hepar
	35	278.8	16.2	1779	22	AAF86101	Human Heparanase-2
	36	270.6	15.7	1685	24	AAD29204	Human heparanase-2
	37	269.4	15.7	1898	22	AAA91098	Human heparanase,
	38	230.6	13.4	385	21	AAA75082	Rat EST which is h
	39	197.6	11.5	1511	24	AAD29205	Human heparanase-2
	40	196.4	11.4	1724	22	AAA91099	Human heparanase,
	41	196.4	11.4	1930	22	AAS13843	DNA encoding novel
	42	196.4	11.4	2300	22	AAH22673	Heparanase-like pr
	43	195.6	11.4	1674	24	AAD29203	Human heparanase-2
	44	195	11.3	3943	22	AAH28347	Nucleotide sequenc
	45	194.4	11.3	1891	24	ABZ11527	Human polynucleoti

ALIGNMENTS

RESULT 1
AAX35648
ID AAX35648 standard; cDNA; 1721 BP.
XX
AC AAX35648;
XX
DT 09-JUL-1999 (first entry)
XX
DE cDNA encoding a human heparanase protein.
XX
KW Heparanase; hpa; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW athersclerosis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW ss.
XX
OS Homo sapiens.
XX
PN WO9911798-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US17954.
XX
PR 02-JUL-1998; 98US-0109386.
PR 02-SEP-1997; 97US-0922170.
XX
PA (FRIE/) FRIEDMAN M M.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Feinstein E, Pecker I, Vlodavsky I;
XX
DR WPI; 1999-302255/25.
DR P-PSDB; AAY02345.
XX
PT New human polynucleotide useful for treating angiogenesis,
PT restenosis, and inflammation
XX
PS Claim 4; Fig 1; 63pp; English.
XX
CC The specification describes a polypeptide having heparanase (hpa)
CC activity. The recombinant protein is used as a modulator of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections
CC or disintegration of neurodegenerative plaques. Heparanase may be
CC useful for conditions such as wound healing, angiogenesis, restenosis,
CC atherscleros is, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC and renal failure in biopsy specimens, plasma samples, and body fluids.
CC The present sequence encodes human heparanase.
XX
SQ Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

Query Match 100.0%; Score 1721; DB 20; Length 1721;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1721; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
```

```
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721


RESULT 2
AAA75051
ID   AAA75051 standard; cDNA; 1721 BP.
XX
AC   AAA75051;
XX
DT   15-JAN-2001  (first entry)
XX
DE   cDNA encoding a human heparanase polypeptide.
XX
KW   Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW   heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW   wound healing; infection; burn; angiogenesis; restenosis;
KW   atherosclerosis; inflammation; neurodegenerative disease;
KW   Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1693
FT                   /*tag=  a
FT                   /product= "heparanase"
FT   stem_loop       698..724
FT                   /*tag=  b
FT                   /note= "these nucleotides are likely to be involved
FT                          in forming stem and loop structures"
XX
PN   WO200052178-A1.
XX
PD   08-SEP-2000.
XX
PF   14-FEB-2000; 2000WO-US03542.
XX
PR   01-MAR-1999;   99US-0258892.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Feinstein E;
XX
DR   WPI; 2000-579289/54.
DR   P-PSDB; AAB08849.
XX
PT   New polynucleotides encoding a polypeptide having heparanase activity,
PT   useful in wound healing and in gene therapy, particularly in treating
PT   tumour, inflammation, autoimmunity, neurodegenerative diseases  -
XX
PS   Claim 9; Fig 1; 152pp; English.
XX
CC   The present sequence encodes a human protein with heparanase catalytic
CC   activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC   particularly in treating tumour, inflammation or autoimmunity.
CC   Particularly, the polynucleotide is useful in modulating the
CC   bioavailability of heparin-binding growth factors, cellular responses
CC   to heparin-binding growth factors (e.g. bFGF) and cytokines
CC   (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
CC   cellular susceptibility to certain viral and some bacterial and protozoa
CC   infections, or disintegration of neurodegenerative plaques. The
CC   polynucleotide is also useful in wound healing (e.g. thermal, chemical
CC   or radiation burns), and in the treatment of angiogenesis, restenosis,
CC   atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC   Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC   bacterial or protozoa infections.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match             100.0%;  Score 1721;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
```

```
Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
             |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 3
AAZ39195
ID   AAZ39195 standard; cDNA; 1721 BP.
XX
AC   AAZ39195;
XX
DT   02-MAR-2000  (first entry)
XX
DE   Human heparanase encoding cDNA.
XX
KW   Human; heparanase; hpa; genetic modification; expression; anticancer;
KW   angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW   anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW   heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW   metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW   neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW   micrometastasis; autoimmune lesion; kidney failure; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1694
FT                   /*tag= a
FT                   /product= "heparanase"
XX
PN   WO9957244-A1.
XX
PD   11-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09256.
XX
PR   01-MAY-1998;   98US-0071618.
PR   02-MAR-1999;   99US-0260038.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Ben-Artzi H,  Ayal-Hershkovitz M,  Yacoby-Zeevi O,  Pecker I,  Peleg Y;
PI   Shlomi Y;
XX
DR   WPI; 2000-062144/05.
DR   P-PSDB; AAY57590.
XX
PT   Engineered cells that express recombinant heparanase, useful
PT   therapeutically, e.g. for treating angiogenesis and to screen for
PT   specific inhibitors, potential anticancer agents -
XX
PS   Claim 2; Page 106-107; 118pp; English.
XX
CC   The present invention describes genetically modified cells (A) containing
CC   a polynucleotide (I) that encodes a polypeptide with heparanase activity,
CC   and express recombinant heparanase (II). Heparanase cleaves heparan
CC   sulphate (HS) at specific intrachain sites, resulting in release of
CC   heparin-binding growth factors, enzymes and proteins that are sequestered
CC   by HS in basement membranes, extracellular matrix or cell surfaces. It
CC   may also be implicated in tumour angiogenesis and metastases. (II) is
CC   potentially useful in wound healing and for treating angiogenesis,
CC   restenosis, atherosclerosis, inflammation, neurodegeneration, viral
CC   infection and cystic fibrosis. It can also be used to neutralise heparin
CC   (an alternative to protamine) and to screen for specific inhibitors
CC   (potentially useful for treating cancer and metastases). Antibodies
CC   raised against (II) are used for immunodetection and diagnosis of
CC   micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
CC   in large quantities, in a form that is homogeneously processed and
CC   activated/neutralised by a dedicated protease. The present sequence
CC   encodes human heparanase.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;
```

```
  Query Match             100.0%;  Score 1721;  DB 21;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
```

```
Qy       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
Db       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
Db       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
Db       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
Db       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
Db       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
Db       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
Db       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
Db       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
Db       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
Db       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
Db      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
Db      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
Db      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
Db      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
Db      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
Db      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
Db      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
Db      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
Db      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
Db      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
Db      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
Db      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

RESULT 4
AAZ33290
ID AAZ33290 standard; cDNA; 1721 BP.
XX
AC AAZ33290;
XX
DT 21-FEB-2000 (first entry)
XX
DE Human heparanase nucleotide sequence.
XX
KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW autoimmune disease; anticancer; kidney disease; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   63..1694
FT   /*tag= a
FT   /product= "heparanase"
XX
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999;   99WO-US09255.
XX
PR 01-MAY-1998;   98US-0071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
DR P-PSDB; AAY52990.
XX
PT Heparanase-specific molecular probes useful for diagnosis and
PT treatment, e.g. of tumors, and for targeted drug delivery -
XX
PS Example; Page 82-84; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and
CC anti-heparanase antibodies are used to detect or quantify the expression
CC of heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary,
CC cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
CC disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
CC syndrome, sepsis and inflammatory or autoimmune disease), for targeted

```
CC   drug delivery (e.g. of anticancer agents) and as research reagents.
CC   The present sequence encodes human heparanase, which is used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

  Query Match           100.0%;  Score 1721;  DB 21;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721


RESULT 5
AAA91112
ID   AAA91112 standard; DNA; 1721 BP.
XX
AC   AAA91112;
XX
DT   20-APR-2001  (first entry)
XX
DE   Human heparanase, coding sequence fragment isolated from EST clone.
XX
KW   Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW   atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW   neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW   gene therapy; mouse; expressed sequence tag; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100643-A2.
XX
```

PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL00358.
XX
PR 25-JUN-1999; 99US-0140801.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI; 2001-137930/14.
XX
PT New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy
PT protocols for angiogenesis, restenosis, atherosclerosis, or
PT inflammation -
XX
PS Example 1; Page 67; 67pp; English.
XX
CC This sequence represents a human heparanase coding sequence clone,
CC isolated from an EST clone. The invention relates to heparanase DNA
CC and protein sequences. The heparanase DNA and protein sequences are
CC useful in wound healing, angiogenesis, restenosis, atherosclerosis,
CC inflammation, pulmonary diseases, neurodegenerative diseases (such as
CC Scrape, Alzheimer's disease, and Creutzfeldt-Jakob disease) or viral
CC infections. The heparanase coding sequence is particularly useful in gene
CC therapy.
XX
SQ Sequence 1721 BP; 451 A; 413 C; 410 G; 447 T; 0 other;

Query Match 100.0%; Score 1721; DB 22; Length 1721;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1721; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
```

```
Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
             ||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

RESULT 6
AAX35650
ID AAX35650 standard; cDNA; 1899 BP.
XX
AC AAX35650;
XX
DT 09-JUL-1999 (first entry)
XX
DE cDNA encoding a human heparanase protein.
XX
KW Heparanase; hpa; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherscleroSis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW ss.
XX
OS Homo sapiens.
XX
PN WO9911798-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US17954.
XX
PR 02-JUL-1998; 98US-0109386.
PR 02-SEP-1997; 97US-0922170.
XX
PA (FRIE/) FRIEDMAN M M.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Feinstein E, Pecker I, Vlodavsky I;
XX
DR WPI; 1999-302255/25.
DR P-PSDB; AAY02346.
XX
PT New human polynucleotide useful for treating angiogenesis,
PT restenosis, and inflammation
XX
PS Claim 4; Page 64-65; 63pp; English.
XX
CC The specification describes a polypeptide having heparanase (hpa)
CC activity. The recombinant protein is used as a modulator of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections
CC or disintegration of neurodegenerative plaques. Heparanase may be
CC useful for conditions such as wound healing, angiogenesis, restenosis,
CC atherscleroSis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC and renal failure in biopsy specimens, plasma samples, and body fluids.
CC The present sequence encodes human heparanase.
XX
SQ Sequence 1899 BP; 495 A; 433 C; 510 G; 461 T; 0 other;

Query Match 99.9%; Score 1719.4; DB 20; Length 1899;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1720; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 60
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 CTAGAGCTCTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 238

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 298

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 358

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 418

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 478

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 538

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 598

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 658

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 718

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 778

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 838

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 898

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 958

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 1018

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 1078

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 1138

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1198

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1258

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1318

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1319 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1378

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1438

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1498

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1558

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1618

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1678

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1738

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1798

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1858

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
             |||||||||||||||||||||||||||||||||||||||||
Db       1859 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1899


RESULT 7
AAA75053
ID   AAA75053 standard; cDNA; 1899 BP.
XX
AC   AAA75053;
XX
DT   15-JAN-2001  (first entry)
XX
DE   cDNA encoding a human heparanase polypeptide.
XX
KW   Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW   heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW   wound healing; infection; burn; angiogenesis; restenosis;
KW   atherosclerosis; inflammation; neurodegenerative disease;
KW   Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             94..1872
FT                   /*tag=  a
FT                   /product= "heparanase"
XX
PN   WO200052178-A1.
XX
PD   08-SEP-2000.
XX
PF   14-FEB-2000; 2000WO-US03542.
XX
PR   01-MAR-1999;   99US-0258892.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA   (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA   (FRIE/) FRIEDMAN M M.
XX
PI   Pecker I,  Vlodavsky I,  Feinstein E;
XX
DR   WPI; 2000-579289/54.
DR   P-PSDB; AAB08850.
XX
PT   New polynucleotides encoding a polypeptide having heparanase activity,
PT   useful in wound healing and in gene therapy, particularly in treating
PT   tumour, inflammation, autoimmunity, neurodegenerative diseases  -
XX
PS   Claim 9; Page 121-122; 152pp; English.
XX
CC   The present sequence encodes a human protein with heparanase catalytic
CC   activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC   particularly in treating tumour, inflammation or autoimmunity.
CC   Particularly, the polynucleotide is useful in modulating the
CC   bioavailability of heparin-binding growth factors, cellular responses
CC   to heparin-binding growth factors (e.g. bFGF) and cytokines
CC   (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
CC   cellular susceptibility to certain viral and some bacterial and protozoa
CC   infections, or disintegration of neurodegenerative plaques. The
CC   polynucleotide is also useful in wound healing (e.g. thermal, chemical
CC   or radiation burns), and in the treatment of angiogenesis, restenosis,
CC   atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC   Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC   bacterial or protozoa infections.
XX
SQ   Sequence 1899 BP; 495 A; 433 C; 510 G; 461 T; 0 other;

  Query Match             99.9%;  Score 1719.4;  DB 21;  Length 1899;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1720;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 CTAGAGCTCTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 238

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 298

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 358

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 418

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 478

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 538

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 598

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 658

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 718

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 778
```

```
Qy       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       779 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 838

Qy       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       839 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 898

Qy       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       899 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 958

Qy       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       959 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 1018

Qy       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1019 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 1078

Qy       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1079 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 1138

Qy       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1139 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1198

Qy      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1199 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1258

Qy      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1259 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1318

Qy      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1319 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1378

Qy      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1379 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1438

Qy      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1439 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1498

Qy      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1499 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1558

Qy      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1559 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1618

Qy      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1619 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1678

Qy      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1679 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1738

Qy      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1739 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1798

Qy      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1799 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1858

Qy      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
            |||||||||||||||||||||||||||||||||||||||||
Db      1859 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1899
```

RESULT 8
AAF93788
ID AAF93788 standard; cDNA; 1722 BP.
XX
AC AAF93788;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human cDNA encoding a membrane or secretory protein clone PSEC0090.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; ss.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
DR P-PSDB; AAB88361.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 1; SEQ ID 89; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 1722 BP; 449 A; 414 C; 412 G; 447 T; 0 other;

Query Match 99.5%; Score 1713; DB 22; Length 1722;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1716; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
```

```
Db          2 CTAGAGCTCTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 61

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Db         62 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 121

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Db        122 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 181

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Db        182 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 241

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Db        242 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 301

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Db        302 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 361

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Db        362 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 421

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Db        422 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 481

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Db        482 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 541

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Db        542 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 601

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Db        602 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAGGAACAGCAGATTTGCAGTGGA 661

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Db        662 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 721

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Db        722 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 781

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Db        782 CGCAGTTAGGAGAAGATTTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 841

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Db        842 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 901

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Db        902 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 961

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Db        962 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1021

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Db       1022 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1081

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Db       1082 GGTTAGGAGAAACAAGCTCTGCACATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1141

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Db       1142 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1201

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Db       1202 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1261

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Db       1262 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1321

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Db       1322 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1381

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Db       1382 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAATG 1441

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Db       1442 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1501

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Db       1502 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1561

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Db       1562 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1621

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Db       1622 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1681

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721

Db       1682 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1722
```

RESULT 9
AAX37259
ID AAX37259 standard; DNA; 1713 BP.
XX
AC AAX37259;
XX
DT 21-JUL-1999 (first entry)
XX
DE Human heparanase enzyme encoding DNA.
XX
KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG; ss.
XX
OS Homo sapiens.
XX
PN WO9921975-A1.
XX
PD 06-MAY-1999.
XX
PF 28-OCT-1998; 98WO-AU00898.
XX
PR 09-DEC-1997; 97AU-0000812.
PR 28-OCT-1997; 97AU-0000062.
XX
PA (AUSU) UNIV AUSTRALIAN NAT.
XX
PI Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;
XX

DR WPI; 1999-312956/26.
DR P-PSDB; AAY17082.
XX
PT Polynucleotides encoding mammalian endoglucuronidases, especially
PT heparanases, useful to promote wound healing
XX
PS Claim 3; Page 69-73; 112pp; English.
XX
CC The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG. The
CC present sequence represents a DNA encoding human heparanase.
XX
SQ Sequence 1713 BP; 460 A; 404 C; 406 G; 443 T; 0 other;

Query Match 98.5%; Score 1694.6; DB 20; Length 1713;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1697; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      18 CCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCG 77
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCG 60

Qy      78 AAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCC 137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCC 120

Qy     138 CCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACC 197
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCTGGTGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACC 180

Qy     198 CAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTG 257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTG 240

Qy     258 GCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGA 317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGA 300

Qy     318 GGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGAT 377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGAT 360

Qy     378 CCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGAT 437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGAT 420

Qy     438 ATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCC 497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCC 480

Qy     498 TACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTAC 557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTAC 540

Qy     558 TCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATC 617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATC 600

Qy     618 TTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAG 677
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAG 660

Qy     678 TTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAA 737
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAA 720

Qy     738 CCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGAT 797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGAT 780

Qy     798 TATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGT 857
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGT 840

Qy     858 CCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCT 917
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCT 900

Qy     918 GGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCT 977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCT 960

Qy     978 ACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAA 1037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAA 1020

Qy    1038 GTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGC 1097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGC 1080

Qy    1098 TCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGG 1157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGG 1140

Qy    1158 CTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTC 1217
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTC 1200

Qy    1218 TTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGG 1277
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGG 1260

Qy    1278 CTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGT 1337
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGT 1320

Qy    1338 TCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTAT 1397
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTAT 1380

Qy    1398 AAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGG 1457
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAATGTCACCAAGTACTTGCGG 1440

Qy    1458 TTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCT 1517
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCT 1500

Qy    1518 CATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGAT 1577
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGAT 1560

Qy    1578 CAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCA 1637
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCA 1620

Qy    1638 GCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAA 1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAA 1680

Qy    1698 TAAAATATACTAGTCCTGACA 1718
           ||||||||||||||||||| |
Db    1681 TAAAATATACTAGTCCTGAAA 1701
```

```
RESULT 10
AAX37260
ID   AAX37260 standard; DNA; 1723 BP.
XX
AC   AAX37260;
XX
DT   21-JUL-1999  (first entry)
XX
DE   Seq ID No: 14 of WO9921975.
XX
KW   Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW   metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW   arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW   human; HSPG; ss.
XX
OS   Homo sapiens.
XX
PN   WO9921975-A1.
XX
PD   06-MAY-1999.
XX
PF   28-OCT-1998;   98WO-AU00898.
XX
PR   09-DEC-1997;   97AU-0000812.
PR   28-OCT-1997;   97AU-0000062.
XX
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Freeman CG,  Hamdorf BJ,  Hulett MD,  Parish CR;
XX
DR   WPI; 1999-312956/26.
DR   P-PSDB; AAY17083.
XX
PT   Polynucleotides encoding mammalian endoglucuronidases, especially
PT   heparanases, useful to promote wound healing
XX
PS   Claim 11; Page 76-79; 112pp; English.
XX
CC   The invention relates to nucleic acid sequences that encode heparanase
CC   enzymes having  endoglucuronidase activity. Recombinant heparanases are
CC   capable of removing the HS side chain from heparan sulfate proteoglycan
CC   (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC   inhibit heparanase, this is useful for treatment of a physiological or
CC   medical condition associated with elevated heparanase activity, such as
CC   metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC   arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC   rat heparanases can be used to enhance wound healing, especially
CC   associated with tissue development and repair. The conditions mentioned
CC   above can be diagnosed using specific antibodies, and also using primers
CC   and probes specific for the heparanase polynucleotides. Other uses of the
CC   heparanases include sequencing sulfated molecules such as HSPG.
XX
SQ   Sequence 1723 BP; 461 A; 407 C; 412 G; 443 T; 0 other;

  Query Match           98.1%;  Score 1688.8;  DB 20;  Length 1723;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches 1701;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         18 CCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCG 77
Db          7 CCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCG 66

Qy         78 AAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCC 137
Db         67 AAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCC 126

Qy        138 CCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACC 197
Db        127 CCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACC 186

Qy        198 CAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTG 257
Db        187 CAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTG 246

Qy        258 GCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGA 317
Db        247 GCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGA 306

Qy        318 GGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGAT 377
Db        307 GGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGAT 366

Qy        378 CCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGAT 437
Db        367 CCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGAT 426

Qy        438 ATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCC 497
Db        427 ATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCC 486

Qy        498 TACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTAC 557
Db        487 TACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTAC 546

Qy        558 TCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATC 617
Db        547 TCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATC 606

Qy        618 TTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAG 677
Db        607 TTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAG 666

Qy        678 TTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAA 737
Db        667 TTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAA 726

Qy        738 CCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGAT 797
Db        727 CCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGAT 786

Qy        798 TATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGT 857
Db        787 TTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGT 846

Qy        858 CCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCT 917
Db        847 CCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCT 906

Qy        918 GGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCT 977
Db        907 GGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCT 966

Qy        978 ACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAA 1037
Db        967 ACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAA 1026

Qy       1038 GTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGC 1097
Db       1027 GTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGC 1086

Qy       1098 TCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGG 1157
Db       1087 TCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGG 1146

Qy       1158 CTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTC 1217
Db       1147 CTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTC 1206

Qy       1218 TTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGG 1277
Db       1207 TTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGG 1266

Qy       1278 CTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGT 1337
Db       1267 CTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGT 1326
```

```
Qy       1338 TCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTAT 1397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1327 TCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTAT 1386

Qy       1398 AAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGG 1457
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1387 AAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAATGTCACCAAGTACTTGCGG 1446

Qy       1458 TTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCT 1517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1447 TTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCT 1506

Qy       1518 CATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGAT 1577
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1507 CATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGAT 1566

Qy       1578 CAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCA 1637
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1567 CAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGG-TTGCCA 1625

Qy       1638 GCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAA 1697
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 GCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAA 1685

Qy       1698 TAAAATATACTAGTCCTGACACTG 1721
             |||||||||||||||||||||||
Db       1686 TAAAATATACTAGTCCTGACACTG 1709
```

```
RESULT 11
AAX86671
ID   AAX86671 standard; cDNA; 3726 BP.
XX
AC   AAX86671;
XX
DT   20-MAR-2003  (updated)
DT   14-OCT-1999  (first entry)
XX
DE   cDNA encoding a human protein with heparanase activity.
XX
KW   Human; heparanase; heparan sulfate; trauma; autoimmune disease;
KW   skin disease; cardiovascular disease; nervous system disease;
KW   Alzheimer's disease; cancer; cancer metastasis; angiogenesis;
KW   inflammation; arthritis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..1768
FT                   /*tag=  a
XX
PN   WO9940207-A1.
XX
PD   12-AUG-1999.
XX
PF   05-FEB-1999;   99WO-EP00777.
XX
PR   09-FEB-1998;   98GB-0002725.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Nakajima M,  Toyoshima M;
XX
DR   WPI; 1999-494300/41.
DR   P-PSDB; AAY30124.
XX
PT   New heparanase polypeptide useful for treating autoimmune diseases,
PT   skin diseases, cardiovascular diseases and nervous system diseases
PT   including Alzheimer's disease
XX
PS   Claim 5; Page 27-29; 40pp; English.
XX
CC   The present sequence encodes a polypeptide with human heparanase
CC   biological activity. Antagonists and inhibitors of the protein prevent
CC   it from degrading the extracellular matrix and releasing heparan sulfate
CC   from the extracellular matrix surface. The heparanase protein or the
CC   anti-heparanase antibody are used in pharmaceutical compositions for
CC   treating warm blooded animals suffering from a disease resulting from
CC   shortage or lack of the heparanase protein, or from excessive activity
CC   or over-expression of the heparanase protein, respectively. The
CC   heparanase protein is used in treating diseases such as trauma,
CC   autoimmune disease, skin diseases, cardiovascular diseases and nervous
CC   system diseases including Alzheimer's disease resulting from shortage or
CC   lack of polypeptide. The anti-heparanase antibody is used in treating
CC   the diseases like cancer, cancer metastasis, angiogenesis and
CC   inflammation including arthritis resulting from excessive activity or
CC   over expression of heparanase protein. The anti-heparanase antibody can
CC   be used to detect the presence or absence of polypeptide and its
CC   concentration.
CC   (Updated on 20-MAR-2003 to correct PA field.)
XX
SQ   Sequence 3726 BP; 1030 A; 802 C; 765 G; 1129 T; 0 other;
```

```
  Query Match             98.0%;  Score 1686.8;  DB 20;  Length 3726;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1688;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         32 CTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCGAAGCCTGCGCTGCC 91
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 CTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCGAAGCCTGCGCTGCC 165

Qy         92 GCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCC 151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 GCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCC 225

Qy        152 CCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCA 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 CCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCA 285

Qy        212 CCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCG 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 CCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCG 345

Qy        272 GTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGC 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 GTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGC 405

Qy        332 GTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATC 391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 GTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATC 465

Qy        392 AACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGG 451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 AACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGG 525

Qy        452 ATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATT 511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 ATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATT 585

Qy        512 GCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGT 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 GCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGT 645

Qy        572 AGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCCTAAATGC 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 AGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCCTAAATGC 705

Qy        632 GTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTA 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 GTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTA 765

Qy        692 CTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCT 751
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        766 CTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCT 825
```

```
Qy        752 TAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTATATTCAATTGCA 811
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        826 TAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTTTATTCAATTGCA 885

Qy        812 TAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCA 871
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        886 TAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCA 945

Qy        872 GCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGAT 931
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        946 GCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGAT 1005

Qy        932 TGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTT 991
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1006 TGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTT 1065

Qy        992 TCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGT 1051
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1066 TCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGT 1125

Qy       1052 TGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGG 1111
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1126 TGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGG 1185

Qy       1112 CGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGG 1171
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1186 CGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGG 1245

Qy       1172 CCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAA 1231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1246 CCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAA 1305

Qy       1232 CTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTT 1291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1306 CTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTT 1365

Qy       1292 CAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAA 1351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 CAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAA 1425

Qy       1352 GCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTT 1411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTT 1485

Qy       1412 AACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGGTTACCCTATCCTTT 1471
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1486 AACTCTGTATGCCATAAACCTCCATAATGTCACCAAGTACTTGCGGTTACCCTATCCTTT 1545

Qy       1472 TTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTC 1531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1546 TTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTC 1605

Qy       1532 CAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACC 1591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1606 CAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACC 1665

Qy       1592 TTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAG 1651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1666 TTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAG 1725

Qy       1652 TTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAATAAAATATACTAGT 1711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1726 TTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAATAAAATATACTAGT 1785

Qy       1712 CCTGACACTG 1721
              ||||||||||
Db       1786 CCTGACACTG 1795
```

RESULT 12
AAH20940
ID AAH20940 standard; cDNA; 1724 BP.
XX
AC AAH20940;
XX
DT 24-AUG-2001 (first entry)
XX
DE Human heparanase inhibitor cDNA.
XX
KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;
KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
KW primary cardiomyopathy; peripheral odema; pulmonary congestion;
KW hepatic congestion; hydrothorax; ascite; nocturia; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 52..1683
FT /*tag= a
FT /product= "heparanase"
XX
PN DE19955803-A1.
XX
PD 23-MAY-2001.
XX
PF 19-NOV-1999; 99DE-1055803.
XX
PR 19-NOV-1999; 99DE-1055803.
XX
PA (KNOL) KNOLL AG.
XX
PI Herr D, Hahn A, Laux V;
XX
DR WPI; 2001-368371/39.
DR P-PSDB; AAB86206.
XX
PT Treatment or prevention of cardiac insufficiency and related
PT conditions, e.g. pulmonary congestion and dyspnoea, comprises
PT administration of heparanase inhibitor -
XX
PS Disclosure; Page 8-11; 16pp; German.
XX
CC This invention describes a novel heparanase inhibitor which can be used
CC for the treatment or prevention of cardiac insufficiency and associated
CC indications, symptoms and/or malfunctions. The heparanase inhibitor of
CC the invention has cardiant, nephrotropic and hepatotropic activity. The
CC products of the invention can be used in human and veterinary medicine,
CC for the treatment or prevention of congestive heart failure e.g. primary
CC cardiomyopathy. Associated conditions treated or prevented with the
CC inhibitor are especially peripheral odemas, pulmonary and hepatic
CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
CC nocturia can also be treated. This sequence encodes the human heparanase
CC protein described in the method of the invention.
XX
SQ Sequence 1724 BP; 466 A; 405 C; 410 G; 443 T; 0 other;

Query Match 97.8%; Score 1682.6; DB 22; Length 1724;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1696; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy         18 CCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCG 77
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db          8 CCGCTGCGCGGCAGCTGGC-GGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCG 66

Qy         78 AAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCC 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 AAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCC 126

Qy        138 CCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACC 197
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CCTGGTGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACC 186

Qy        198 CAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 CAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTG 246
```

```
Qy       258 GCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGA 317
Db       247 GCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGA 306

Qy       318 GGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGAT 377
Db       307 GGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGAT 366

Qy       378 CCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGAT 437
Db       367 CCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGAT 426

Qy       438 ATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCC 497
Db       427 ATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCC 486

Qy       498 TACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTAC 557
Db       487 TACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTAC 546

Qy       558 TCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATC 617
Db       547 TCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATC 606

Qy       618 TTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAG 677
Db       607 TTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAG 666

Qy       678 TTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAA 737
Db       667 TTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAA 726

Qy       738 CCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGAT 797
Db       727 CCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGAT 786

Qy       798 TATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGT 857
Db       787 TTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGT 846

Qy       858 CCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCT 917
Db       847 CCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCT 906

Qy       918 GGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCT 977
Db       907 GGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCT 966

Qy       978 ACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAA 1037
Db       967 ACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAA 1026

Qy      1038 GTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGC 1097
Db      1027 GTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGC 1086

Qy      1098 TCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGG 1157
Db      1087 TCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGG 1146

Qy      1158 CTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTC 1217
Db      1147 CTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTC 1206

Qy      1218 TTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGG 1277
Db      1207 TTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGG 1266

Qy      1278 CTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGT 1337
Db      1267 CTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGT 1326

Qy      1338 TCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTAT 1397
Db      1327 TCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTAT 1386

Qy      1398 AAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGG 1457
Db      1387 AAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAATGTCACCAAGTACTTGCGG 1446

Qy      1458 TTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCT 1517
Db      1447 TTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCT 1506

Qy      1518 CATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGAT 1577
Db      1507 CATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGAT 1566

Qy      1578 CAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCA 1637
Db      1567 CAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCA 1626

Qy      1638 GCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAA 1697
Db      1627 GCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAA 1686

Qy      1698 TAAAATATACTAGTCCTGACA 1718
Db      1687 TAAAATATACTAGTCCTGAAA 1707
```

RESULT 13
ABZ22816
ID ABZ22816 standard; cDNA; 1669 BP.
XX
AC ABZ22816;
XX
DT 02-APR-2003 (first entry)
XX
DE Human heparanase encoding cDNA SEQ ID NO:17.
XX
KW Human; heparanase; phosphorothioate; antisense oligonucleotide;
KW cytostatic; gene therapy; tumour; enzyme; gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1..1638
FT   /*tag= a
FT   /product= "heparanase"
XX
PN WO2003004705-A1.
XX
PD 16-JAN-2003.
XX
PF 01-JUL-2002; 2002WO-US20636.
XX
PR 05-JUL-2001; 2001US-0899440.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Stein C;
XX
DR WPI; 2003-201558/19.
DR P-PSDB; ABP56822.
XX
PT New oligonucleotide having a sequence complementary to a sequence of
PT ribonucleic acid encoding a heparanase, useful for preparing a
PT composition for treating tumor -
XX
PS Disclosure; Fig 3; 48pp; English.
XX
CC The present invention describes an oligonucleotide having a sequence
CC complementary to a sequence of ribonucleic acid encoding a heparanase.
CC The oligonucleotide hybridises with the ribonucleic acid under conditions
CC of high stringency and has a sequence comprising 10-40 bp. The

```
CC   internucleoside linkages of the oligonucleotide comprise at least one
CC   phosphorothioate linkage. Hybridisation of the oligonucleotide to the
CC   ribonucleic acid inhibits expression of the heparanase, where inhibition
CC   of heparanase means at least a 50% reduction in the quality of
CC   heparanase. Also described: (1) a method of inhibiting expression of a
CC   heparanase in a cell; (2) a composition comprising the above
CC   oligonucleotide in an amount effective to inhibit the expression of
CC   heparanase in the cell and a carrier; and (3) a method of treating a
CC   tumour in a subject comprises administering to the subject an amount of
CC   the above oligonucleotide effective to inhibit expression of a heparanase
CC   in the subject. Heparanase antisense oligonucleotides have cytostatic
CC   activity, can be used in gene therapy, and can be used for preparing a
CC   composition for treating tumours. The present sequence encodes human
CC   heparanase, which is given in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1669 BP; 445 A; 396 C; 388 G; 440 T; 0 other;

  Query Match             94.8%;  Score 1631.4;  DB 25;  Length 1669;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1658;  Conservative    0;  Mismatches    1;  Indels    6;  Gaps    2;

Qy         63 ATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCC---GCTGATGCTGCTGCTCCTGGGG 119
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db          1 ATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGCTGATGCTGCTGCTCCTGGGG 60

Qy        120 CCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCA---CAGGACGTC 176
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db         61 CCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGCAGGACGTC 120

Qy        177 GTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCC 236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCC 180

Qy        237 GTCACCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCA 296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCACCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCA 240

Qy        297 AAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAG 356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAG 300

Qy        357 ACAGACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGG 416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACAGACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGG 360

Qy        417 CAATCTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAG 476
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATCTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAG 420

Qy        477 AAGTTACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAA 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAGTTACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAA 480

Qy        537 AAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAAC 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAAC 540

Qy        597 TGCTCAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAG 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGCTCAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAG 600

Qy        657 TGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATT 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATT 660

Qy        717 TCTTGGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAAT 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCTTGGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAAT 720

Qy        777 GGGTCGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTC 836
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        721 GGGTCGCAGTTAGGAGAAGATTTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTC 780

Qy        837 AAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATG 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATG 840

Qy        897 CTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTAC 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTAC 900

Qy        957 TATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATT 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATT 960

Qy       1017 TTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAG 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAG 1020

Qy       1077 GTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACC 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACC 1080

Qy       1137 TTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAA 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAA 1140

Qy       1197 GTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTC 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTC 1200

Qy       1257 GATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTG 1316
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTG 1260

Qy       1317 TTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACA 1376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACA 1320

Qy       1377 AACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCAT 1436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCAT 1380

Qy       1437 AACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATAC 1496
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATAC 1440

Qy       1497 CTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTA 1556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTA 1500

Qy       1557 ACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCA 1616
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCA 1560

Qy       1617 GGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAA 1676
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAA 1620

Qy       1677 GTTGCTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTGCTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1665


RESULT 14
AAZ11236
ID   AAZ11236 standard; cDNA; 1593 BP.
XX
AC   AAZ11236;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Human pre-proheparanase coding sequence.
XX
KW   Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
```

KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;
KW inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
KW tumour growth; fibroproliferative disorder; neurodegenerative disease;
KW therapy; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1593
FT /*tag= a
FT /product= pre-proheparanase
XX
PN WO9943830-A2.
XX
PD 02-SEP-1999.
XX
PF 18-FEB-1999; 99WO-US01489.
XX
PR 26-MAR-1998; 98US-0079401.
PR 24-FEB-1998; 98US-0075706.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Fairbanks MB, Heinrikson RL, Mildner AM;
XX
DR WPI; 1999-540598/45.
DR P-PSDB; AAY34173.
XX
PT New isolated platelet heparanase polypeptides, used to develop
PT products for, e.g. wound healing and blocking angiogenesis
XX
PS Claim 2; Fig 7; 57pp; English.
XX
CC This sequence encodes the human pre-proheparanase of the invention. This
CC sequence was isolated from human platelets. The heparanase can be used
CC for identifying agents which alter heparanase activity. The heparanase
CC can be used for wound healing or for blocking angiogenesis or
CC inflammation. It can be used for treating e.g. psoriasis, diabetic
CC retinopathy or solid tumours, or for the degradation of heparin and the
CC neutralisation of heparin's anticoagulant properties during surgery.
CC Inhibitors of heparanase activity can be used in the treatment of
CC arthritis, asthma, and other inflammatory diseases, vascular restenosis,
CC atherosclerosis, tumour growth and progression, fibroproliferative
CC disorders, and central nervous system (CNS) and neurodegenerative
CC diseases. The products can also be used for detection and diagnosis. The
CC purified heparanase, both recombinantly produced human heparanase and
CC heparanase isolated from human platelet activity, allows for the
CC convenient selection of compounds having anti-heparanase activity,
CC i.e. inhibitors of heparanase activity, by measuring inhibition of
CC heparanase activity. Inhibition of heparanase activity can be measured by
CC blocking heparanase-mediated release of radioactive fragments from in
CC vivo radiolabelled (HSPG)/heparin.
XX
SQ Sequence 1593 BP; 426 A; 370 C; 369 G; 428 T; 0 other;

Query Match 92.1%; Score 1585; DB 20; Length 1593;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1588; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      102 ATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCG 161
            ||||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||
Db        1 ATGCTGCTGCTCCTGGGGCCGCTGGGTCCCTTCTCCCCTGGCGCCTTGCCCCGACCTGCG 60

Qy      162 CAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGC 221
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGC 120

Qy      222 CCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATC 281
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATC 180

Qy      282 CTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGG 341
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGG 240

Qy      342 TTTGGTGGCACCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAA 401
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TTTGGTGGCACCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAA 300

Qy      402 GAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCT 461
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCT 360

Qy      462 CCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGA 521
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGA 420

Qy      522 GAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTA 581
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTA 480

Qy      582 TACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGA 641
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 TACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGA 540

Qy      642 ACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCC 701
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 ACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCC 600

Qy      702 AAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCT 761
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 AAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCT 660

Qy      762 GATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTA 821
            ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      661 GATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTTTATTCAATTGCATAAACTTCTA 720

Qy      822 AGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGA 881
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 AGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGA 780

Qy      882 AAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTT 941
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 AAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTT 840

Qy      942 ACATGGCATCACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCT 1001
            ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      841 ACATGGCATCACTACTATTTGAATGGACGGACTGCTACCAAGGAAGATTTTCTAAACCCT 900

Qy     1002 GATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACC 1061
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 GATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACC 960

Qy     1062 AGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCC 1121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 AGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCC 1020

Qy     1122 TTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCC 1181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 TTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCC 1080

Qy     1182 CGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTA 1241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 CGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTA 1140

Qy     1242 GTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTG 1301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 GTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTG 1200

Qy     1302 GTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTA 1361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 GTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTA 1260

Qy     1362 TACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTAT 1421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1261 TACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTAT 1320

Qy       1422 GCCATAAACCTCCATAACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAG 1481
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCCATAAACCTCCATAATGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAG 1380

Qy       1482 CAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTC 1541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTC 1440

Qy       1542 CAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAA 1601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAA 1500

Qy       1602 AAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTG 1661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTG 1560

Qy       1662 ATAAGAAATGCCAAAGTTGCTGCTTGCATCTGA 1694
              |||||||||||||||||||||||||||||||||
Db       1561 ATAAGAAATGCCAAAGTTGCTGCTTGCATCTGA 1593
```

```
RESULT 15
ABL40753
ID   ABL40753 standard; cDNA; 1584 BP.
XX
AC   ABL40753;
XX
DT   03-JUL-2002  (first entry)
XX
DE   Chicken signal peptide/human heparanase chimeric cDNA.
XX
KW   Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
KW   anti-protozoan; neuroprotective; heparin; chicken; human; chimeric; ss.
XX
OS   Synthetic.
OS   Gallus gallus.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1584
FT                   /*tag=  a
FT                   /product= "chimeric chicken-human heparanase"
FT   sig_peptide     1..57
FT                   /*tag=  b
FT                   /note= "chicken heparanase signal peptide"
FT   mat_peptide     58..1581
FT                   /*tag=  c
FT                   /note= "human mature heparanase"
XX
PN   US2002034810-A1.
XX
PD   21-MAR-2002.
XX
PF   16-AUG-2001; 2001US-0930218.
XX
PR   20-SEP-2000; 2000US-0666390.
XX
PA   (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI   Goldshmidt O,  Pecker I,  Vlodavsky I,  Michal I,  Zcharia E;
XX
DR   WPI; 2002-338926/37.
DR   P-PSDB; ABB07815.
XX
PT   Nucleic acid encoding avian and reptile heparanase polypeptide is
PT   useful to treat various heparin-related disorders and the signal
PT   peptide is useful in production of membrane-targeted or secreted
PT   recombinant proteins  -
XX
PS   Disclosure; Page 24-25; 39pp; English.
XX
CC   The invention relates to an isolated avian and reptile nucleic acid,
CC   encoding a polypeptide with heparanase catalytic activity. The signal
CC   peptide of the nucleic acid can be used to express membrane-associated or
CC   secreted proteins in heterologous expression systems. The encoded
CC   polypeptides can be used to prevent tumour angiogenesis, metastasis and
CC   invasion, and to intervene with pathologies associated with impaired
CC   heparin-binding growth factors, cellular responses to heparin-binding
CC   growth factors and cytokines, cell interaction with plasma lipoproteins,
CC   cellular susceptibility to viral, protozoa and bacterial infections or
CC   disintegration of neurodegenerative plaques. The present sequence
CC   represents a chicken signal peptide/human heparanase chimeric cDNA
CC   sequence.
XX
SQ   Sequence 1584 BP; 424 A; 361 C; 373 G; 426 T; 0 other;
```

```
  Query Match           89.2%;  Score 1535;  DB 24;  Length 1584;
  Best Local Similarity  98.1%;  Pred. No. 0;
  Matches 1553;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy        112 TCCTGGGGCCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGG 171
              | |||| || ||||   | | |    |   |||  |||| |       |   |||||||
Db          2 TGCTGGTGCTGCTGCTGCTCGTGCTGCTGCTCGCTGTGCCGCCGAGGCGGACGGCACAGG 61

Qy        172 ACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCC 231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 ACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCC 121

Qy        232 TGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTT 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTT 181

Qy        292 CTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCA 241

Qy        352 CCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTT 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTT 301

Qy        412 ACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGG 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 ACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGG 361

Qy        472 AGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACC 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACC 421

Qy        532 AGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTG 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 AGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTG 481

Qy        592 CAAACTGCTCAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATT 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 CAAACTGCTCAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATT 541

Qy        652 TGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATA 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATA 601

Qy        712 ACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCA 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 ACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCA 661

Qy        772 TCAATGGGTCGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCA 831
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        662 TCAATGGGTCGCAGTTAGGAGAAGATTTTATTCAATTGCATAAACTTCTAAGAAAGTCCA 721

Qy        832 CCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTA 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 CCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTA 781

Qy        892 AGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        782 AGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATC 841

Qy        952 ACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGG 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 ACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGG 901

Qy       1012 ACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCA 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 ACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCA 961

Qy       1072 AGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 AGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCG 1021

Qy       1132 ACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAA 1191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 ACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAA 1081

Qy       1192 TAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAA 1251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 TAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAA 1141

Qy       1252 ACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCA 1311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 ACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCA 1201

Qy       1312 AGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATT 1371
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 AGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATT 1261

Qy       1372 GCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACC 1431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1262 GCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACC 1321

Qy       1432 TCCATAACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATA 1491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 TCCATAACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATA 1381

Qy       1492 AATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATG 1551
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 AATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATG 1441

Qy       1552 GTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCC 1611
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCC 1501

Qy       1612 GGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATG 1671
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1502 GGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATG 1561

Qy       1672 CCAAAGTTGCTGCTTGCATCTGA 1694
              |||||||||||||||||||||||
Db       1562 CCAAAGTTGCTGCTTGCATCTGA 1584
```

Search completed: February 16, 2004, 09:18:09
Job time : 850.407 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 08:49:50 ; Search time 6685.2 Seconds
(without alignments)
6256.802 Million cell updates/sec

Title: US-10-676-079-1
Perfect score: 1721
Sequence: 1 ctagagctttcgactctccg.....atatactagtcctgacactg 1721

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1093.6	63.5	2173	11	AK040471	AK040471 Mus muscu
c	2	954.2	55.4	1156	9	AL552151	AL552151 AL552151
	3	933.4	54.2	1962	11	AK087283	AK087283 Mus muscu
	4	923.6	53.7	1185	9	AL552174	AL552174 AL552174
	5	902.6	52.4	1201	9	AL545270	AL545270 AL545270
c	6	866.4	50.3	1200	9	AL545232	AL545232 AL545232
c	7	844.8	49.1	1083	13	BX398409	BX398409 BX398409
c	8	804.6	46.8	914	13	BX373611	BX373611 BX373611
	9	761.6	44.3	881	14	CB988510	CB988510 AGENCOURT
	10	750.2	43.6	924	13	BQ691142	BQ691142 AGENCOURT
	11	689.8	40.1	907	13	BQ438834	BQ438834 AGENCOURT
c	12	687.8	40.0	708	13	BQ775819	BQ775819 UI-H-FH0-
	13	615.2	35.7	652	13	BX107633	BX107633 BX107633
c	14	614.8	35.7	682	12	BM996417	BM996417 UI-H-DT0-
c	15	553.4	32.2	556	9	AI342512	AI342512 qt28g01.x
c	16	541	31.4	549	10	BF197674	BF197674 7m96b03.x
c	17	537.4	31.2	540	9	AI582254	AI582254 tq65f01.x
c	18	486.8	28.3	553	14	N30824	N30824 yw70a03.s1
c	19	486.4	28.3	495	9	AI660639	AI660639 wf23a12.x
c	20	470.4	27.3	484	9	AI033490	AI033490 ox06c03.s
c	21	468.2	27.2	479	9	AI824984	AI824984 wb04f06.x
c	22	466.2	27.1	488	13	BU617228	BU617228 UI-H-DF0-
	23	465.8	27.1	587	14	N45367	N45367 yw97a02.r1
c	24	456.6	26.5	518	14	N30845	N30845 yw70d08.s1
	25	426.4	24.8	629	14	CB483444	CB483444 jns07_E02
c	26	410.2	23.8	427	9	AI928508	AI928508 wp02g09.x
c	27	404.8	23.5	412	9	AI027968	AI027968 ov84d05.x
c	28	371.2	21.6	480	9	AW465382	AW465382 BP230019A
c	29	369.6	21.5	386	9	AI393542	AI393542 tg46a08.x
	30	352	20.5	370	14	N41349	N41349 yw70a03.r1
c	31	348.4	20.2	367	9	AW015299	AW015299 UI-H-BI0-
c	32	341	19.8	445	10	BF040685	BF040685 BP250002A
	33	332	19.3	484	14	N41373	N41373 yw70d08.r1
	34	327.2	19.0	632	10	BB636167	BB636167 BB636167
	35	320.4	18.6	506	10	BF419897	BF419897 UI-R-BJ2-
c	36	317.4	18.4	341	9	AI189657	AI189657 qd18h01.x
c	37	314.4	18.3	327	9	AA948627	AA948627 oq37c01.s
c	38	310.4	18.0	408	14	N32056	N32056 yw97a02.s1
	39	304.6	17.7	431	9	AA177901	AA177901 mt05b03.r
	40	304.4	17.7	453	9	AW429437	AW429437 67351 MAR
	41	299	17.4	451	9	AA674378	AA674378 vp99c03.r
	42	271.2	15.8	447	10	BB815166	BB815166 BB815166
	43	263.4	15.3	452	10	BG663158	BG663158 DRA1A11 R
c	44	254.4	14.8	283	9	AI207288	AI207288 qg09h01.x
c	45	253.8	14.7	777	9	AL718139	AL718139 AL718139

ALIGNMENTS

RESULT 1
AK040471
LOCUS AK040471 2173 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430101M04 product:heparanase, full insert sequence.
ACCESSION AK040471
VERSION AK040471.1 GI:26333764
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2173)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1. .2173
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A430101M04"
/db_xref="taxon:10090"
/clone="A430101M04"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
CDS 91. .1698
/note="unnamed protein product; heparanase (MGD|MGI:1343124, GB|AA047943, evidence: BLASTN, 99%, match=480)
putative"
/codon_start=1
/protein_id="BAC30600.1"
/db_xref="GI:26333765"
/translation="MLRLLLLWLWGPLGALAQGAPAGTAPTDDVVDLEFYTKRPLRSV
SPSFLSITIDASLATDPRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKEP
TSEERSYWKSQVNHDICRSEPVSAAVLRKLQVEWPFQELLLLREQYQKEFKNSTYSRS
SVDMLYSFAKCSGLDLIFGLNALLRTPDLRWNSSNAQLLLDYCSSKGYNISWELGNEP
NSFWKKAHILIDGLQLGEDFVELHKLLQRSAFQNAKLYGPDIGQPRGKTVKLLRSFLK
AGGEVIDSLTWHHYYLNGRIATKEDFLSSDVLDTFILSVQKILKVTKEITPGKKVWLG
ETSSAYGGGAPLLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNYHLVDENFEP
LPDYWLSLLFKKLVGPRVLLSRVKGPDRSKLRVYLHCTNVYHPRYQEGDLTLYVLNLH
NVTKHLKVPPPLFRKPVDTYLLKPSGPDGLLSKSVQLNGQILKMVDEQTLPALTEKPL
PAGSALSLPAFSYGFFVIRNAKIAACI"
BASE COUNT 532 a 536 c 564 g 541 t
ORIGIN

Query Match 63.5%; Score 1093.6; DB 11; Length 2173;
Best Local Similarity 79.7%; Pred. No. 5.5e-291;
Matches 1291; Conservative 0; Mismatches 329; Indels 0; Gaps 0;

```
Qy        83 TGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCCCCTGG 142
             || | ||| |  || |||| |||||  ||||  ||||||||| ||| ||||  |||  ||
Db        87 TGGGATGCTGAGGCTGCTGCTGCTGTGGCTCTGGGGGCCGCTCGGTGCCCTGGCCCAGGG 146

Qy       143 CGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAGGA 202
             |||||   |  |  | ||||   |  | ||||| || ||| |||| || | |||| ||
Db       147 CGCCCCCGCGGGGACCGCGCCGACCGACGACGTGGTAGACTTGGAGTTTTACACCAAGCG 206

Qy       203 GCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTGGCCAC 262
             ||||||  |     ||||| |||||||||||||||| |||||||  |||||||  |||||||||
Db       207 GCCGCTCCGAAGCGTGAGTCCCTCGTTCCTGTCCATCACCATCGACGCCAGCCTGGCCAC 266

Qy       263 GGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGAGGCTT 322
             ||||||||  |||||||  | |||||||  |||||||  |||  |||  |   ||||  ||||||||
Db       267 CGACCCGCGCTTCCTCACCTTCCTGGGCTCTCCAAGGCTCCGTGCTCTGGCTAGAGGCTT 326

Qy       323 GTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGATCCCAA 382
             ||||||||  |||  ||||  |||||  |||||  |||||  ||||||||  |||||  |||||   |
Db       327 ATCTCCTGCATACTTGAGATTTGGCGGCACAAAGACTGACTTCCTTATTTTTGATCCGGA 386

Qy       383 GAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTG 442
             ||||||  |  ||  |  |||||  ||||||||||||  ||||||||||||||||  ||||||||
Db       387 CAAGGAACCGACTTCCGAAGAAAGAAGTTACTGGAAATCTCAAGTCAACCATGATATTTG 446

Qy       443 CAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCA 502
             ||  | ||   |  || || |||  |||  |  |||  | | | ||||||||||||  |||
Db       447 CAGGTCTGAGCCGGTCTCTGCTGCGGTGTTGAGGAAACTCCAGGTGGAATGGCCCTTCCA 506

Qy       503 GGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAG 562
             |||||  |||||  ||||||||  || |||||  ||  ||||||||||||||||||||||||
Db       507 GGAGCTGTTGCTGCTCCGAGAGCAGTACCAAAAGGAGTTCAAGAACAGCACCTACTCAAG 566

Qy       563 AAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGG 622
             ||||||  ||  ||   ||||  ||||  ||||||  ||  |||||  ||   |  |||  ||||||||||
Db       567 AAGCTCAGTGGACATGCTCTACAGTTTTGCCAAGTGCTCGGGGTTAGACCTGATCTTTGG 626

Qy       623 CCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCT 682
             ||||||||||||  ||  ||||  ||||  ||  |  |||||||||  ||  ||  ||  |||  |  ||
Db       627 TCTAAATGCGTTACTACGAACCCCAGACTTACGGTGGAACAGCTCCAACGCCCAGCTTCT 686
```

```
Qy       683 CCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAA 742
Db       687 CCTTGACTACTGCTCTTCCAAGGGTTATAACATCTCCTGGGAACTGGGCAATGAGCCCAA 746

Qy       743 CAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTATAT 802
Db       747 CAGTTTCTGGAAGAAAGCTCACATTCTCATCGATGGGTTGCAGTTAGGAGAAGACTTTGT 806

Qy       803 TCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGA 862
Db       807 GGAGTTGCATAAACTTCTACAAAGGTCAGCTTTCCAAAATGCAAAACTCTATGGTCCTGA 866

Qy       863 TGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGG 922
Db       867 CATCGGTCAGCCTCGAGGGAAGACAGTTAAACTGCTGAGGAGTTTCCTGAAGGCTGGCGG 926

Qy       923 AGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCTACCAG 982
Db       927 AGAAGTGATCGACTCTCTTACATGGCATCACTATTACTTGAATGGACGCATCGCTACCAA 986

Qy       983 GGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTT 1042
Db       987 AGAAGATTTTCTGAGCTCTGATGTGCTGGACACTTTTATTCTCTCTGTGCAAAAAATTCT 1046

Qy      1043 CCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGC 1102
Db      1047 GAAGGTCACTAAAGAGATCACACCTGGCAAGAAGGTCTGGTTGGGAGAGACGAGCTCAGC 1106

Qy      1103 ATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGA 1162
Db      1107 TTACGGTGGCGGTGCACCCTTGCTGTCCAACACCTTTGCAGCTGGCTTTATGTGGCTGGA 1166

Qy      1163 TAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGG 1222
Db      1167 TAAATTGGGCCTGTCAGCCCAGATGGGCATAGAAGTCGTGATGAGGCAGGTGTTCTTCGG 1226

Qy      1223 AGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATC 1282
Db      1227 AGCAGGCAACTACCACTTAGTGGATGAAAACTTTGAGCCTTTACCTGATTACTGGCTCTC 1286

Qy      1283 TCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAA 1342
Db      1287 TCTTCTGTTCAAGAAACTGGTAGGTCCCAGGGTGTTACTGTCAAGAGTGAAAGGCCCAGA 1346

Qy      1343 GAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGA 1402
Db      1347 CAGGAGCAAACTCCGAGTGTATCTCCACTGCACTAACGTCTATCACCCACGATATCAGGA 1406

Qy      1403 AGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGGTTACC 1462
Db      1407 AGGAGATCTAACTCTGTATGTCCTGAACCTCCATAATGTCACCAAGCACTTGAAGGTACC 1466

Qy      1463 CTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGG 1522
Db      1467 GCCTCCGTTGTTCAGGAAACCAGTGGATACGTACCTTCTGAAGCCTTCGGGGCCGGATGG 1526

Qy      1523 ATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAAC 1582
Db      1527 ATTACTTTCCAAATCTGTCCAACTGAACGGTCAAATTCTGAAGATGGTGGATGAGCAGAC 1586

Qy      1583 CTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTT 1642
Db      1587 CCTGCCAGCTTTGACAGAAAAACCTCTCCCCGCAGGAAGTGCACTAAGCCTGCCTGCCTT 1646

Qy      1643 CTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAATAAAA 1702
Db      1647 TTCCTATGGTTTTTTTGTCATAAGAAATGCCAAAATCGCTGCTTGTATATGAAAATAAAA 1706
```

RESULT 2
AL552151/c
LOCUS AL552151 1156 bp mRNA linear EST 31-MAY-2003

```
DEFINITION  AL552151 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI059YN15 3-PRIME, mRNA sequence.
ACCESSION   AL552151
VERSION     AL552151.2  GI:31273967
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12890775.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI059CG08NP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI059CG08NP1.
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI059YN15"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      305 a    251 c    233 g    323 t     44 others
ORIGIN
```

Query Match 55.4%; Score 954.2; DB 9; Length 1156;
Best Local Similarity 98.0%; Pred. No. 1.5e-252;
Matches 970; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

```
Qy       698 TTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAA 757
Db       989 YTTCCAAGGGTAWTACATTTCTTGGRACTAGGCAAATGAACCTAACAG-TTCCTTAAGAA 931

Qy       758 GGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTATATTCAATTGCATAAACT 817
Db       930 GGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTTTATTCAATTGCATAAACT 871

Qy       818 TCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCG 877
Db       870 TCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCG 811

Qy       878 AAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTC 937
Db       810 AAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTC 751

Qy       938 AGTTACATGGCATCACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAA 997
Db       750 AGTTACATGGCATCACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAA 691

Qy       998 CCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAG 1057
Db       690 CCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAG 631

Qy      1058 CACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGC 1117
Db       630 CACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGC 571
```

```
Qy       1118 GCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTC 1177
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db        570 GCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCWGTC 511

Qy       1178 AGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCA 1237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         510 AGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCA 451

Qy       1238 TTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAA 1297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         450 TTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAA 391

Qy       1298 ATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCG 1357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         390 ATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCG 331

Qy       1358 AGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCT 1417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         330 AGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCT 271

Qy       1418 GTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAA 1477
             |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||||
Db         270 GTATGCCATAAACCTCCATAATGTCACCAAGCACTTGCGGTTACCCTATCCTTTTTCTAA 211

Qy       1478 CAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATC 1537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         210 CAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATC 151

Qy       1538 TGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAAT 1597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         150 TGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAAT 91

Qy       1598 GGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTT 1657
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          90 GGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTNTCTCATATAGTTTTTT 31

Qy       1658 TGTGATAAGAAATGCCAAAGTTGCTGCTTG 1687
             ||||||||||||||||||||||||||||:|
Db          30 TGTGATAAGAAATGCCAAAGTTGCTGCTKG 1

RESULT 3
AK087283
LOCUS       AK087283               1962 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 0 day neonate lung cDNA, RIKEN full-length enriched
            library, clone:E030042C09 product:heparanase, full insert sequence.
ACCESSION   AK087283
VERSION     AK087283.1  GI:26104170
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1962)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .1962
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:E030042C09"
                     /db_xref="taxon:10090"
```

```
                     /clone="E030042C09"
                     /tissue_type="lung"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="0 day neonate"
     misc_feature    112. .1568
                     /note="heparanase (MGD|MGI:1343124, GB|AA047943, evidence:
                     BLASTN, 99%, match=480)
                     putative"
BASE COUNT      480 a    487 c    480 g    515 t
ORIGIN

  Query Match           54.2%;  Score 933.4;  DB 11;  Length 1962;
  Best Local Similarity   78.7%;  Pred. No. 1.2e-246;
  Matches 1152;  Conservative    0;  Mismatches  306;  Indels    5;  Gaps    3;

Qy         83 TGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCCCCTGG 142
              || | ||| |  || |||| |||||  |||| |||||||||| ||| ||||  |||  ||
Db        108 TGGGATGCTGAGGCTGCTGCTGCTGTGGCTCTGGGGGCCGCTCGGTGCCCTGGCCCAGGG 167

Qy        143 -CGCCCTGCCCCCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAG- 200
               |||||    |  |  | ||||    |  | ||||| || ||| |||| || | |||| ||
Db        168 CCGCCCCCGCGGGGACCGCGCCGACCGACGACGTGGTAGACTTGGAGTTTTACACCAAGC 227

Qy        201 GAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACC---TG 257
              |  ||||| |      ||||| |||||||||||||| ||||||| |||||| | |     |
Db        228 GGCCCGCTCCGAAGCGTGAGTCCCTCGTTCCTGTCCATCACCATCGACGCCCAGCCTGGG 287

Qy        258 GCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGA 317
              |||| |||||||| ||||||| ||||||| ||| |  |||| |||
Db        288 CCCACCGACCCGCGCTTCCTCACCTTCCTGGGCTCTCCAAGGCTCCGTGCTCTGGCTAGA 347

Qy        318 GGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGAT 377
              ||||| |||||||| ||| |||| ||||| ||||| ||||| |||||||| ||||| |||
Db        348 GGCTTATCTCCTGCATACTTGAGATTTGGCGGCACAAAGACTGACTTCCTTATTTTTGAT 407

Qy        378 CCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGAT 437
              ||  | ||||||  | ||  |  ||||| |||||||||||| |||||||||||||| |||
Db        408 CCGGACAAGGAACCGACTTCCGAAGAAAGAAGTTACTGGAAATCTCAAGTCAACCATGAT 467

Qy        438 ATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCC 497
              |||||||  | ||    |  || || |||  |||  |  |||  | | | ||||||||||
Db        468 ATTTGCAGGTCTGAGCCGGTCTCTGCTGCGGTGTTGAGGAAACTCCAGGTGGAATGGCCC 527

Qy        498 TACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTAC 557
              | ||||||||  ||||| |||||||| || ||||| ||  ||||||||||||||||||||
Db        528 TTCCAGGAGCTGTTGCTGCTCCGAGAGCAGTACCAAAAGGAGTTCAAGAACAGCACCTAC 587

Qy        558 TCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATC 617
              |||||||||| || ||  |||| |||| |||||| || ||||| ||  | ||| |||||
Db        588 TCAAGAAGCTCAGTGGACATGCTCTACAGTTTTGCCAAGTGCTCGGGGTTAGACCTGATC 647

Qy        618 TTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAG 677
              |||||  |||||||||||| || ||||  |||| || | |||||||||| || || || |||
Db        648 TTTGGTCTAAATGCGTTACTACGAACCCCAGACTTACGGTGGAACAGCTCCAACGCCCAG 707

Qy        678 TTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAA 737
              | ||||| ||||||||||||||||||||| |||||||| || |||||||| ||||||||
Db        708 CTTCTCCTTGACTACTGCTCTTCCAAGGGTTATAACATCTCCTGGGAACTGGGCAATGAG 767

Qy        738 CCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGAT 797
              || |||||||||   ||||| ||| | ||| ||||| |||||| ||||||||||||||||
Db        768 CCCAACAGTTTCTGGAAGAAAGCTCACATTCTCATCGATGGGTTGCAGTTAGGAGAAGAC 827

Qy        798 TATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGT 857
              | | |  |||||||||||||||  || |||  | ||| |||||||||||||||||||||
Db        828 TTTGTGGAGTTGCATAAACTTCTACAAAGGTCAGCTTTCCAAAATGCAAAACTCTATGGT 887

Qy        858 CCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCT 917
              |||||  | |||||||||||| | ||||| | |||  |||||| ||| ||||||||||||
Db        888 CCTGACATCGGTCAGCCTCGAGGGAAGACAGTTAAACTGCTGAGGAGTTTCCTGAAGGCT 947

Qy        918 GGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGACGGACTGCT 977
              || |||||||||| || ||  |||||||||||||||| || |||||||||| |  |||

Db        948 GGCGGAGAAGTGATCGACTCTCTTACATGGCATCACTATTACTTGAATGGACGCATCGCT 1007

Qy        978 ACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAA 1037
              ||||  |||||||||||| | | |||||||  |||||| |||||||    ||||||||||||
Db       1008 ACCAAAGAAGATTTTCTGAGCTCTGATGTGCTGGACACTTTTATTCTCTCTGTGCAAAAA 1067

Qy       1038 GTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGC 1097
              || |  ||||    | |     | ||  |||||||||||||||||||| ||||| || |||
Db       1068 ATTCTGAAGGTCACTAAAGAGATCACACCTGGCAAGAAGGTCTGGTTGGGAGAGACGAGC 1127

Qy       1098 TCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGG 1157
              || || || || ||||| || |||||||| ||| ||||||||||||||||||||||||||
Db       1128 TCAGCTTACGGTGGCGGTGCACCCTTGCTGTCCAACACCTTTGCAGCTGGCTTTATGTGG 1187

Qy       1158 CTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTC 1217
              |||||||||||||||||||||||||  ||||| |||||||| |||||||||| || |||
Db       1188 CTGGATAAATTGGGCCTGTCAGCCCAGATGGGCATAGAAGTCGTGATGAGGCAGGTGTTC 1247

Qy       1218 TTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGG 1277
              || |||||||| |||||||| |||||||||||||||| || |||||||||||||| |||
Db       1248 TTCGGAGCAGGCAACTACCACTTAGTGGATGAAAACTTTGAGCCTTTACCTGATTACTGG 1307

Qy       1278 CTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGT 1337
              || |||||||||||||||||| |||| ||  ||| ||||||| || |||| ||| ||||
Db       1308 CTCTCTCTTCTGTTCAAGAAACTGGTAGGTCCCAGGGTGTTACTGTCAAGAGTGAAAGGC 1367

Qy       1338 TCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTAT 1397
              || | || || || || ||||| || || || ||||| |||   |  | ||| | |||
Db       1368 CCAGACAGGAGCAAACTCCGAGTGTATCTCCACTGCACTAACGTCTATCACCCACGATAT 1427

Qy       1398 AAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGG 1457
              | |||||||| |||||||||||| | | |||||||||| |||||||| ||||| |
Db       1428 CAGGAAGGAGATCTAACTCTGTATGTCCTGAACCTCCATAATGTCACCAAGCACTTGAAG 1487

Qy       1458 TTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCT 1517
              ||||   ||| || |  |  || | |||||||  |||||||| |  |||| ||| ||
Db       1488 GTACCGCCTCCGTTGTTCAGGAAACCAGTGGATACGTACCTTCTGAAGCCTTCGGGGCCG 1547

Qy       1518 CATGGATTACTTTCCAAATCTGT 1540
              |||||||||||||||| |  ||
Db       1548 GATGGATTACTTTCCAAGTGAGT 1570


RESULT 4
AL552174
LOCUS       AL552174               1185 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL552174 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI059YN15 5-PRIME, mRNA sequence.
ACCESSION   AL552174
VERSION     AL552174.2  GI:31273990
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1185)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12890820.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI059CG08QP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI059CG08QP1.
FEATURES             Location/Qualifiers
```

```
     source          1. .1185
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI059YN15"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      280 a    293 c    293 g    280 t     39 others
ORIGIN

  Query Match             53.7%;  Score 923.6;  DB 9;  Length 1185;
  Best Local Similarity   98.4%;  Pred. No. 4.6e-244;
  Matches  956;  Conservative    7;  Mismatches    6;  Indels    3;  Gaps    3;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||  ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         68 CTAGAGCTCTCGACTCTCCGCTGCGC-GCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 126

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 186

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 246

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 306

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 366

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 426

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 486

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 546

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 606

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 666

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 726

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 786

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 846

Qy        781 CGCAGTTAGGAGAAGATTTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        847 CGCAGTTAGGAGAAGATTTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 906

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTG- 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        907 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 966

Qy        900 AAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTAT 959
              ||||||||||||||||||||||||||||||||||||:|    :||||||| :|:|||
Db        967 AAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAKTACATWGGCATCA-YAMTAT 1025

Qy        960 TTGAATGGACGG 971
              ||||||||::|:
Db       1026 TTGAATGGMSGR 1037


RESULT 5
AL545270
LOCUS       AL545270                1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL545270 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI028YF04 5-PRIME, mRNA sequence.
ACCESSION   AL545270
VERSION     AL545270.2  GI:31267106
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12877751.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI028DC02QP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI028DC02QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI028YF04"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      292 a    282 c    305 g    279 t     43 others
ORIGIN

  Query Match             52.4%;  Score 902.6;  DB 9;  Length 1201;
  Best Local Similarity   99.3%;  Pred. No. 3e-238;
  Matches  916;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy          4 GAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGA 63
              | | | |||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         63 GGGATCTCGACTCTCCGCTGCGCGNCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGA 122

Qy         64 TGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGC 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGCCGC 182

Qy        124 TGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGGACC 242
```

```
Qy        184 TGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCACCA 302

Qy        244 TTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGCTTC 362

Qy        304 GTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAGACT 422

Qy        364 TCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTC 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTC 482

Qy        424 AAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTAC 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 AAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTAC 542

Qy        484 GGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCA 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCA 602

Qy        544 AGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAG 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 AGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAG 662

Qy        604 GACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACA 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 GACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACA 722

Qy        664 GTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGG 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 GTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGG 782

Qy        724 AACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGC 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 AACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGC 842

Qy        784 AGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATG 843
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        843 AGTTAGGAGAAGATTTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATG 902

Qy        844 CAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGA 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 CAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGA 962

Qy        904 GCTTCCTGAAGGCTGGTGGAGA 925
              |||| ||||||||||||||||
Db        963 GCTT-CTGAAGGCTGGTGGAGA 983


RESULT 6
AL545232/c
LOCUS       AL545232               1200 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL545232 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI028YF04 3-PRIME, mRNA sequence.
ACCESSION   AL545232
VERSION     AL545232.2  GI:31267068
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12877713.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI028DC02NP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI028DC02NP1.
FEATURES              Location/Qualifiers
     source           1. .1200
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI028YF04"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      300 a    249 c    249 g    332 t     70 others
ORIGIN

  Query Match           50.3%;  Score 866.4;  DB 9;  Length 1200;
  Best Local Similarity   97.3%;  Pred. No. 3.2e-228;
  Matches  870;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        765 ATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGA 824
              || ||     | |||||||||||||:||||||| ||||||||:||||||||||||||||
Db        894 ATATTTCATCAATGGTCGCAGTTAGGRGAAGATTTTATTCAATKGCATAAACTTCTAAGA 835

Qy        825 AAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAG 884
              ||||||||||||||||:: ||||||||||||||||||||||||||||||||||||||||
Db        834 AAGTCCACCTTCAAAAAWRAAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAG 775

Qy        885 ACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACA 944
              |||||:|| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 ACGGCWAAAATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACA 715

Qy        945 TGGCATCACTACTATTTGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGAT 1004
              ||||||||||||||||:|||||||||||||||||||||||||:  ||||||||||||
Db        714 TGGCATCACTACTATTTGWATGGACGGACTGCTACCAGGGAAGATWAACTAAACCCTGAT 655

Qy       1005 GTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGG 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        654 GTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGG 595

Qy       1065 CCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTG 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        594 CCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTG 535

Qy       1125 CTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGA 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 CTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGA 475

Qy       1185 ATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTG 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 ATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTG 415

Qy       1245 GATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 GATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTG 355

Qy       1305 GGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATAC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 GGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATAC 295

Qy       1365 CTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCC 1424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 CTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCC 235
```

```
Qy       1425 ATAAACCTCCATAACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAA 1484
              ||||||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db        234 ATAAACCTCCATAATGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTTACAAGCAA 175

Qy       1485 GTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAA 1544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        174 GTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAA 115

Qy       1545 CTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAA 1604
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 CTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAA 55

Qy       1605 CCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTT 1658
              ||||||||||||||||||||||||||||||||||||||||||:||:|||||
Db         54 CCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATARTTWTTTT 1


RESULT 7
BX398409/c
LOCUS       BX398409               1083 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX398409 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI058YI24 3-PRIME, mRNA sequence.
ACCESSION   BX398409
VERSION     BX398409.1  GI:30617572
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1083)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI058BE12NP1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI058BE12NP1.
FEATURES             Location/Qualifiers
     source          1. .1083
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI058YI24"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      273 a    264 c    206 g    258 t     82 others
ORIGIN

  Query Match             49.1%;  Score 844.8;  DB 13;  Length 1083;
  Best Local Similarity   89.6%;  Pred. No. 2.9e-222;
  Matches  928;  Conservative   18;  Mismatches   75;  Indels   15;  Gaps    6;

Qy        670 ATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAG 729
              |||||||  ||||||| |||:||||||| |||:||||||||||||||||||||||||||
Db       1045 ATGCTCAAGTGCTCCT-GACYACTGCTC-TCCMAGGGGTATAACATTTCTTGGGAACTAG 988

Qy        730 GCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAG 789
              ||||||||||||||||||:|||||||||||||||| ||||||||||||||||||||||||
Db    .    987 GCAATGAACCTAACAGTTYCCTTAAGAAGGCTGATA-TTTCATCAATGGGTCGCAGTTAG 929

Qy        790 GAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAAC 849
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        928 GAGAAGATTTTATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAAC 869

Qy        850 TCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCC 909
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        868 TCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTRAGATGCTGAAGAGCTTCC 809

Qy        910 TGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGAC 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        808 TGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATTTGAATGGAC 749

Qy        970 GGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTG 1029
              |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        748 GSACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTG 689

Qy       1030 TGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAG 1089
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        688 TGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAG 629

Qy       1090 AAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCT 1149
              |||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db        628 AAACAAGCTCTGCATATGGAGGCGGGGCGCCCTTGCTATCCGACACCTTTGCGGCTGGCT 569

Qy       1150 TTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGC 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        568 TTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGC 509

Qy       1210 AAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTG 1269
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||| ||||
Db        508 AAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGNNNNCTTCGNTCCTTTNCCTG 449

Qy       1270 ATTA--------TTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAAT 1321
              ||          |||| :   : ::   :| : ::|||||||||||||||||||||
Db        448 NTTNGKKGGGKGGTGGCGKKGGGGKGKSGRGGRGGRGKKGGTGGGCACCAAGGTGTTAAT 389

Qy       1322 GGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACAC 1381
              |||||||||||||||||||||||||||||||||| |    ||  |   || |   | |
Db        388 GGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGNGNNNNCCNNCNNNGCNCNNNCNC 329

Qy       1382 TGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGT 1441
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        328 TGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAATGT 269

Qy       1442 CACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCT 1501
              ||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        268 CACCAAGGACTTGCGGTTACCCTATCCTTTTTCTNACAAGCAAGTGGATAAATACCTTCT 209

Qy       1502 AAGACC---TTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAAC 1558
                :  |   || ||| |  |||||:|:||||||    | ||||| | | | |||| | ||
Db        208 GGRGACCGTTTGGGGCCTGCATGGRTKACTTTCNNNNTNTGTCCNANTNANTGGTNTNAC 149

Qy       1559 TCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGG 1618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 TCTAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGG 89

Qy       1619 AAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAG-AAATGCCAAAG 1677
              |||||||||||||||||||||| ||||||||||||||||||||||   |   || ||||||
Db         88 AAGTTCACTGGGCTTGCCAGCTNTCTCATATAGTTTTTTTGTGATNNGNNNATCCCAAAG 29

Qy       1678 TTGCTGCTTGCATCTG 1693
              ||  ||||||:|  :
Db         28 NTGNCGCTTGCMTNNB 13


RESULT 8
BX373611/c
LOCUS       BX373611               914 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX373611 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI058YI24 3-PRIME, mRNA sequence.
ACCESSION   BX373611
VERSION     BX373611.1  GI:30446136
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2469.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAK060DC08NM1&cluster=2469.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAK060DC08NM1.
FEATURES              Location/Qualifiers
     source           1. .914
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI058YI24"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      260 a    191 c    188 g    275 t
ORIGIN

  Query Match             46.8%;  Score 804.6;  DB 13;  Length 914;
  Best Local Similarity   95.3%;  Pred. No. 3.6e-211;
  Matches  872;  Conservative    0;  Mismatches   39;  Indels    4;  Gaps    4;

Qy        806 ATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATGT 865
              ||||||||||||  ||| |||   | ||||||||||||| ||||| |||||||||  |||
Db        913 ATTGCATAAACTCTTAAAAAAAGTCCCCTTCAAAAATGC-AAACTTTATGGTCCTATTGT 855

Qy        866 TGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAGA 925
              |||||| |||| |||||| |||||||| || || ||    ||||| ||||||||||||||
Db        854 TGGTCAACCTCAAAGAAAAACGGCTAAAATTCTAAAAGCTTTCCTAAAGGCTGGTGGAGA 795

Qy        926 AGTGA-TTGATTCAGTTACATGGCATCACTACTATTTGAATGG-ACGGACTGCTACCAGG 983
              | ||| || || ||||   ||| ||||||| |||||||||||| ||||||||||||||||
Db        794 AATGATTTAATCCAGTACAATGCCATCACTTCTATTTGAATGGAACGGACTGCTACCAGG 735

Qy        984 GAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTGTGCAAAAAGTTTTC 1043
              ||||||||  |||||||||||||||||||| |||||||||||||||||||||  |||||
Db        734 GAAGATTTCTTAAACCCTGATGTATTGGACTTTTTTATTTCATCTGTGCAAAAGTTTTTC 675

Qy       1044 CAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCA 1103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 CAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCA 615

Qy       1104 TATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGAT 1163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        614 TATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGAT 555

Qy       1164 AAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGA 1223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        554 AAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGA 495

Qy       1224 GCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCT 1283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        494 GCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCT 435

Qy       1284 CTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAG 1343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 CTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTCAAAG 375

Qy       1344 AGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAA 1403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 AGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAA 315

Qy       1404 GGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGGTTACCC 1463
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        314 GGAGATTTAACTCTGTATGCCATAAACCTCCATAATGTCACCAAGTACTTGCGGTTACCC 255

Qy       1464 TATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGA 1523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 TATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGA 195

Qy       1524 TTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACC 1583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 TTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACC 135

Qy       1584 TTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTC 1643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 TTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTC 75

Qy       1644 TCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAATAAAAT 1703
              |||||||||||||||| ||||||||||||||||||||||||||||| ||||||||||
Db          74 TCATATAGTTTTTTTGT-ATAAGAAATGCCAAAGTTGCTGCTTGCATCTTAAAATAAAAT 16

Qy       1704 ATACTAGTCCTGACA 1718
              |||| ||||| || |
Db          15 ATACCAGTCCCGAAA 1


RESULT 9
CB988510
LOCUS       CB988510                881 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  AGENCOURT_13905817 NIH_MGC_147 Homo sapiens cDNA clone
            IMAGE:30340461 5', mRNA sequence.
ACCESSION   CB988510
VERSION     CB988510.1  GI:30283030
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Stefan Hansson
             cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
            and advice from Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM370  row: f  column: 22
            High quality sequence stop: 664.
FEATURES              Location/Qualifiers
     source           1. .881
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:30340461"
                      /tissue_type="Human Placenta"
                      /lab_host="DH10B TonA"
                      /clone_lib="NIH_MGC_147"
                      /note="Organ: placenta; Vector: pBluescriptR; Site_1:
                      alI-XhoI; Site_2: BamH; Oligo-dT primed using primer
                      5'-TTTTTTTTTTTTTTTTVN-3', size-selected  for average
                      insert size 2.3 kb and normalized to ROT 5. This is a
                      primary library enriched for full-length clones and
                      constructed using  the Cap-trapper method (Carninci, in
```

preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: This is a NIH_MGC library."
BASE COUNT 200 a 244 c 229 g 208 t
ORIGIN

Query Match 44.3%; Score 761.6; DB 14; Length 881;
Best Local Similarity 97.5%; Pred. No. 2.8e-199;
Matches 784; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 CTAGAGCTCTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 119

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 179

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 239

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 299

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 359

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 419

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 479

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 539

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 599

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 659

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 719

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||| |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        720 ACATTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCCAGGGGTATAACATTTCTT 779

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTT-CCTTAAGAAGGCTGATATTTTCATCAATGGG 779
              |||||||||||||||| |||||||||| |||||||||||||||||||||||| |||||
Db        780 GGGAACTAGGCAATGAAACTAACAGTTTCCCTTAAGAAGGCTGATATTTTCATCCATGGG 839

Qy        780 TCGCAGTTAGGAGAAGATTATATT 803
                  |   || |||  || ||
Db        840 GTCGCCATTTTAGGAGAAGATTTT 863
```

RESULT 10
BQ691142
LOCUS BQ691142 924 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8343629 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250265 5', mRNA sequence.
ACCESSION BQ691142
VERSION BQ691142.1 GI:21816458
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 924)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2393 row: a column: 18
High quality sequence stop: 710.
FEATURES Location/Qualifiers
source 1. .924
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6250265"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_110"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 203 a 271 c 227 g 219 t 4 others
ORIGIN

Query Match 43.6%; Score 750.2; DB 13; Length 924;
Best Local Similarity 95.4%; Pred. No. 4.2e-196;
Matches 781; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

```
Qy         32 CTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCGAAGCCTGCGCTGCC 91
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 CTGGCGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCTGCGCTCGAAGCCTGCGCTGCC 163

Qy         92 GCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCC 151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GCCGCCGCTGATGCTGCTGCTCCTGGGGCCGCTGGGTCCCCTCTCCCCTGGCGCCCTGCC 223

Qy        152 CCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCA 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 CCGACCTGCGCAAGCACAGGACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCA 283

Qy        212 CCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCG 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 CCTGGTGAGCCCCTCGTTCCTGTCCGTCACCATTGACGCCAACCTGGCCACGGACCCGCG 343

Qy        272 GTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGC 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GTTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTTGGCCAGAGGCTTGTCTCCTGC 403

Qy        332 GTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATC 391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 GTACCTGAGGTTTGGTGGCACCAAGACAGACTTCCTAATTTTCGATCCCAAGAAGGAATC 463

Qy        392 AACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGG 451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        464 AACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCAAATATGG 523
```

```
Qy        452 ATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATT 511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        524 ATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGGAGCAATT 583

Qy        512 GCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGT 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        584 GCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAAGCTCTGT 643

Qy        572 AGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCCTAAATGC 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        644 AGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCCTANATGC 703

Qy        632 GTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTA 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        704 GTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCCTGGACTA 763

Qy        692 CTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACAGTTTCCT 751
              |||||||| | ||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db        764 CTGCTCTTNCCAGGGGTATAACATTTCTTGGGAACTANGCAATGAACCT-ACAGTTTCCT 822

Qy        752 TAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTATATTCAATTGCA 811
              |||||||      ||       |   |    | ||||||||||| |||| ||||||
Db        823 TAAGAAGCTGATATTTCATCAATGGGTCCCAAGNTAAGGAGAAGATTTTATTTCATTGCA 882

Qy        812 TAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACT 850
              |||||||||||||||||||||||||||||||    ||  ||
Db        883 TAAACTTCTAAGAAAGTCCACCTTCAAAATGCAAACTCT 921
```

RESULT 11
BQ438834

LOCUS	BQ438834 907 bp mRNA linear EST 24-MAY-2002
DEFINITION	AGENCOURT_7761619 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6017952 5', mRNA sequence.
ACCESSION	BQ438834
VERSION	BQ438834.1 GI:21177910
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 907)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM13218 row: b column: 01 High quality sequence stop: 616.

FEATURES Location/Qualifiers

```
     source          1. .907
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6017952"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_70"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.1 kb. Library constructed by Life
                     Technologies."
BASE COUNT      260 a    176 c    226 g    242 t      3 others
ORIGIN

  Query Match             40.1%;  Score 689.8;  DB 13;  Length 907;
  Best Local Similarity   94.8%;  Pred. No. 2.2e-179;
  Matches  778;  Conservative    0;  Mismatches   34;  Indels    9;  Gaps    6;

Qy        385 AGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCA 444
              | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 AAGGATCAACCTTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGATATTTGCA 64

Qy        445 AATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGG 504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AATATGGATCCATCCCTCCTGATGTGGAGGAGAAGTTACGGTTGGAATGGCCCTACCAGG 124

Qy        505 AGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAA 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 AGCAATTGCTACTCCGAGAACACTACCAGAAAAAGTTCAAGAACAGCACCTACTCAAGAA 184

Qy        565 GCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCC 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GCTCTGTAGATGTGCTATACACTTTTGCAAACTGCTCAGGACTGGACTTGATCTTTGGCC 244

Qy        625 TAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCC 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGAACAGTTCTAATGCTCAGTTGCTCC 304

Qy        685 TGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 TGGACTACTGCTCTTCCAAGGGGTATAACATTTCTTGGGAACTAGGCAATGAACCTAACA 364

Qy        745 GTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTATATTC 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        365 GTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGTCGCAGTTAGGAGAAGATTTTATTC 424

Qy        805 AATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATG 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 AATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAAATGCAAAACTCTATGGTCCTGATG 484

Qy        865 TTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAG 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 TTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGAAGAGCTTCCTGAAGGCTGGTGGAG 544

Qy        925 AAGTGATTGATTCAGTTACATGGCATCACTACTAT-TTGAATGGACGGACTGCTACCA-G 982
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
Db        545 AAGTGATTGATTCAGTTACATGGCATCACTACTATNTTGAATGGACGGACTGCTACCAGG 604

Qy        983 GGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTATTTCATCTG--TGCAAAAAGTT 1040
              |||||||||||| |||||||||||||||||||||||||||||||     |||| ||
Db        605 GGAAGATTTTCTANACCCTGATGTATTGGACATTTTTATTTCATCTGGTGCAAAAAGTTT 664

Qy       1041 TTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAG-AAGGTCTGGTTAGGAGAAACAAGCTC 1099
              |||  |||||||||||||||||||||||||||| ||||||||| |  |   |||||||||
Db        665 TTCAGGGTGGTTGAGAGCACCAGGCCTGGCAAGAAAGGTCTGGGTTAGGAGAACAAGCTC 724

Qy       1100 TGCATAT-GGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGC 1158
              ||||||| |||||| |||||||||||||||||||||||||||||||||||||||||     ||
Db        725 TGCATATGGGAGGCNGAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTTATGTGC 784

Qy       1159 T---GGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAA 1196
              |   |   |||||||||||||   || ||||||   | ||
Db        785 TTGGGATAAAATTGGGCCTGTCCAGCCCAATGGGGGAATAA 825
```

RESULT 12
BQ775819/c

LOCUS	BQ775819 708 bp mRNA linear EST 26-JUL-2002
DEFINITION	UI-H-FH0-bcg-a-07-0-UI.s1 NCI_CGAP_FH0 Homo sapiens cDNA clone UI-H-FH0-bcg-a-07-0-UI 3', mRNA sequence.
ACCESSION	BQ775819
VERSION	BQ775819.1 GI:21984295
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 708)

```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FH0-bcg-a-07-0-UI"
                     /tissue_type="Human Chondrosarcoma Cell Line"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FH0"
                     /note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_FH0 is a cDNA library containing the following
                     tissue(s): Human Grade 1 Chondrosarcoma Cell Line The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site. Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is AGAATCCGGC. The cell line was provided by Dr
                     James Martin from University of Iowa
                     TAG_LIB=UI-H-FH0
                     TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
                     Chondrosarcoma
                     TAG_SEQ=AGAATCCGGC"
BASE COUNT      190 a    157 c    144 g    215 t      2 others
ORIGIN

  Query Match             40.0%;  Score 687.8;  DB 13;  Length 708;
  Best Local Similarity   99.4%;  Pred. No. 6.8e-179;
  Matches  689;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1026 TCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTA 1085
             ||||||||||||||| |  ||||||||||||||||||||||||||||||||||||||||||
Db        708 TCTGTGCAAAAAGTNTCNCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCTGGTTA 649

Qy       1086 GGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCT 1145
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 GGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTGCAGCT 589

Qy       1146 GGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATG 1205
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 GGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATG 529

Qy       1206 AGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTA 1265
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 AGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTA 469

Qy       1266 CCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCA 1325
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 CCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCA 409

Qy       1326 AGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGAC 1385
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 AGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGAC 349

Qy       1386 AATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACC 1445
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        348 AATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAATGTCACC 289

Qy       1446 AAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGA 1505
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 AAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGA 229

Qy       1506 CCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAG 1565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 CCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAG 169

Qy       1566 ATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCA 1625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 ATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCA 109

Qy       1626 CTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCT 1685
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        108 CTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCT 49

Qy       1686 TGCATCTGAAAATAAAATATACTAGTCCTGACA 1718
             |||||||||||||||||||||||||||||||||
Db          48 TGCATCTGAAAATAAAATATACTAGTCCTGACA 16


RESULT 13
BX107633
LOCUS       BX107633                652 bp    mRNA    linear   EST 06-FEB-2003
DEFINITION  BX107633 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA
            clone IMAGp998B16561 ; IMAGE:257583, mRNA sequence.
ACCESSION   BX107633
VERSION     BX107633.1  GI:27847484
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
  TITLE     Human UnigeneSet - RZPD3
  JOURNAL   Unpublished
COMMENT     Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998B16561.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Human UnigeneSet - RZPD3 (RZPDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998B16561 ; IMAGE:257583"
                     /dev_stage="two placentae: one from 8 weeks and another
                     from 9 weeks post conception"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'],
```

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 181 a 138 c 149 g 182 t 2 others
ORIGIN

Query Match 35.7%; Score 615.2; DB 13; Length 652;
Best Local Similarity 98.9%; Pred. No. 8.4e-159;
Matches 639; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```
Qy       1016 TTTTATTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAA 1075
              ||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db             7 TTTTTATTTATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAA 66

Qy       1076 GGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACAC 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            67 GGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACAC 126

Qy       1136 CTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGA 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           127 CTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGA 186

Qy       1196 AGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTT 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           187 AGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTT 246

Qy       1256 CGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGT 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           247 CGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGT 306

Qy       1316 GTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCAC 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           307 GTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCAC 366

Qy       1376 AAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCA 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           367 AAACACTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCA 426

Qy       1436 TAACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATA 1495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           427 TAACGTCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATA 486

Qy       1496 CCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCT 1555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           487 CCTTCTAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCT 546

Qy       1556 AACTCTAAAGATGGTGGATGATCAAA-CCTTGCCACCTTTAATGGAAAAA-CCTCTCCGG 1613
              |||||||||||||||||||||||||| |||||||||||||||||||||| |||||||||
Db           547 AACTCTAAAGATGGTGGATGATCAAACCCTTGCCACCTTTAATGGAAAAACCCTCTCCGG 606

Qy       1614 CCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTG 1659
              |||||||||||||||||||||||||| |||||||||||||||  ||
Db           607 CCAGGAAGTTCACTGGGCTTGCCAGCTNTCTCATATAGTTTTTNTG 652
```

RESULT 14
BM996417/c
LOCUS BM996417 682 bp mRNA linear EST 17-JUN-2002
DEFINITION UI-H-DT0-avl-m-03-0-UI.s1 NCI_CGAP_DT0 Homo sapiens cDNA clone IMAGE:5881130 3', mRNA sequence.
ACCESSION BM996417
VERSION BM996417.1 GI:19721318
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 682)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .682
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5881130"
/tissue_type="Metastatic Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_DT0"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_DT0 is a cDNA library containing the following tissue(s): Metatastic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTGTTCGG.
TAG_LIB=UI-H-DT0
TAG_TISSUE=lung metatastic chondrosarcoma
TAG_SEQ=AACTGTTCGG"

BASE COUNT 182 a 146 c 139 g 210 t 5 others
ORIGIN

Query Match 35.7%; Score 614.8; DB 12; Length 682;
Best Local Similarity 98.8%; Pred. No. 1.1e-158;
Matches 660; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

```
Qy       1054 AGAGCACCAGGCCTGGCAAGAAGGTCTGGTTAGGAGAAACAAGCTCTGCATATGGAGGCG 1113
              |||||||||| |||||||||| ||||| ||||||||||||||||||||||||||||||||
Db           682 AGAGCACCAGNCCTGGCAAGAA-GTCTGNTTAGGAGAAACAAGCTCTGCATATGGAGGCG 624

Qy       1114 GAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCC 1173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           623 GAGCGCCCTTGCTATCCGACACCTTTGCAGCTGGCTTTATGTGGCTGGATAAATTGGGCC 564

Qy       1174 TGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACT 1233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           563 TGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTTTGGAGCAGGAAACT 504

Qy       1234 ACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCA 1293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           503 ACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCTATCTCTTCTGTTCA 444

Qy       1294 -AGAAATTGGTGGGCACCAA-GGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAA 1351
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db           443 NAGAAATTGGTGGGCACCAANGGTGTTAATGGCAAGCGTGCAAGGTTCAAAGAGAAGGAA 384

Qy       1352 GCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTT 1411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           383 GCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAAAGAAGGAGATTT 324

Qy       1412 AACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGGTTACCCTATCCTTT 1471
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db           323 AACTCTGTATGCCATAAACCTCCATAATGTCACCAAGTACTTGCGGTTACCCTATCCTTT 264
```

```
Qy       1472 TTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTC 1531
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 TTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCATGGATTACTTTC 204

Qy       1532 CAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACC 1591
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCAAACCTTGCCACC 144

Qy       1592 TTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAG 1651
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 TTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGCTTTCTCATATAG 84

Qy       1652 TTTTTTTGTGATAAGAAATGCC-AAAGTTGCTGCTTGCATCTGAAAATAAAATATACTAG 1710
             ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          83 TTTTTTTGTGATAAGAAATGCCNAAAGTTGCTGCTTGCATCTGAAAATAAAATATACTAG 24

Qy       1711 TCCTGACA 1718
             |||||| |
Db          23 TCCTGAAA 16


RESULT 15
AI342512/c
LOCUS       AI342512                556 bp    mRNA    linear   EST 13-FEB-1999
DEFINITION  qt28g01.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1949328 3', mRNA sequence.
ACCESSION   AI342512
VERSION     AI342512.1  GI:4079439
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 912   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1949328"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      159 a    116 c    113 g    167 t      1 others
ORIGIN

  Query Match           32.2%;  Score 553.4;  DB 9;  Length 556;
  Best Local Similarity   99.6%;  Pred. No. 1e-141;
  Matches  554;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1160 GGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTT 1219
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        556 GGATANATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGGTGATGAGGCAAGTATTCTT 497

Qy       1220 TGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCT 1279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 TGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATCCTTTACCTGATTATTGGCT 437

Qy       1280 ATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTC 1339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 ATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAATGGCAAGCGTGCAAGGTTC 377

Qy       1340 AAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAA 1399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACACTGACAATCCAAGGTATAA 317

Qy       1400 AGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACGTCACCAAGTACTTGCGGTT 1459
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        316 AGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAATGTCACCAAGTACTTGCGGTT 257

Qy       1460 ACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCA 1519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTCTAAGACCTTTGGGACCTCA 197

Qy       1520 TGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCA 1579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTCTAAAGATGGTGGATGATCA 137

Qy       1580 AACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGC 1639
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 AACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAAGTTCACTGGGCTTGCCAGC 77

Qy       1640 TTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAATA 1699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          76 TTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATCTGAAAATA 17

Qy       1700 AAATATACTAGTCCTG 1715
              ||||||||||||||||
Db          16 AAATATACTAGTCCTG 1


Search completed: February 16, 2004, 13:40:49
Job time : 6691.2 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 07:56:25 ; Search time 8216.88 Seconds
(without alignments)
8568.399 Million cell updates/sec

Title: US-10-676-079-1
Perfect score: 1721
Sequence: 1 ctagagctttcgactctccg.....atatactagtcctgacactg 1721

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1721	100.0	1721	6	AR080679	AR080679 Sequence
2	1721	100.0	1721	6	AR080680	AR080680 Sequence
3	1721	100.0	1721	6	AR125603	AR125603 Sequence
4	1721	100.0	1721	6	AR125604	AR125604 Sequence
5	1721	100.0	1721	6	AR194189	AR194189 Sequence
6	1721	100.0	1721	6	AR194190	AR194190 Sequence
7	1721	100.0	1721	6	AR221285	AR221285 Sequence
8	1721	100.0	1721	6	AR221286	AR221286 Sequence
9	1721	100.0	1721	6	AR243203	AR243203 Sequence
10	1721	100.0	1721	6	AR243204	AR243204 Sequence
11	1721	100.0	1721	6	AR287435	AR287435 Sequence
12	1721	100.0	1721	6	AR287436	AR287436 Sequence
13	1721	100.0	1721	6	BD074427	BD074427 Polynucle
14	1721	100.0	1721	6	BD074428	BD074428 Polynucle
15	1719.4	99.9	1899	6	BD074430	BD074430 Polynucle
16	1719.4	99.9	1899	6	BD074431	BD074431 Polynucle
17	1717.8	99.8	1758	9	AF144325	AF144325 Homo sapi
18	1713	99.5	1722	6	AX136167	AX136167 Sequence
19	1713	99.5	1722	6	BD123536	BD123536 Secretory
20	1713	99.5	1722	9	AK075400	AK075400 Homo sapi
21	1694.6	98.5	1713	6	AR156691	AR156691 Sequence
22	1694.6	98.5	1713	6	AX034643	AX034643 Sequence
23	1688.8	98.1	1723	6	AR156692	AR156692 Sequence
24	1688.8	98.1	1723	6	AX034645	AX034645 Sequence
25	1686.8	98.0	3726	6	AR235866	AR235866 Sequence
26	1686.8	98.0	3726	6	AX019348	AX019348 Sequence
27	1686.8	98.0	3726	9	BD131218	BD131218 Human hep
28	1686.8	98.0	3726	9	AF155510	AF155510 Homo sapi
29	1683.8	97.8	1810	9	BC051321	BC051321 Homo sapi
30	1682.6	97.8	1724	6	AX147946	AX147946 Sequence
31	1682.6	97.8	1724	9	AF165154	AF165154 Homo sapi
32	1660.8	96.5	1694	9	AF152376	AF152376 Homo sapi
33	1631.4	94.8	1669	9	AF084467	AF084467 Homo sapi
34	1585	92.1	1593	6	AR210040	AR210040 Sequence
35	1585	92.1	1593	6	BD136761	BD136761 Human pla
36	1254.4	72.9	1662	4	AF281160	AF281160 Bos tauru
37	1099.2	63.9	1736	10	AY077467	AY077467 Mus muscu
38	1081.6	62.8	3177	10	AF184967	AF184967 Rattus no
39	802.2	46.6	1380	6	AR156693	AR156693 Sequence
40	802.2	46.6	1380	6	AX034647	AX034647 Sequence
41	786.2	45.7	1191	6	AR156694	AR156694 Sequence
42	786.2	45.7	1191	6	AX034649	AX034649 Sequence
43	595	34.6	1605	5	AY037007	AY037007 Gallus ga
c 44	486.8	28.3	553	6	AX336600	AX336600 Sequence
45	453	26.3	824	6	AR080681	AR080681 Sequence

ALIGNMENTS

RESULT 1
AR080679
LOCUS AR080679 1721 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 9 from patent US 5968822.
ACCESSION AR080679
VERSION AR080679.1 GI:10007409
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1721)
AUTHORS Pecker,I., Vlodavsky,I. and Feinstein,E.
TITLE Polynucleotide encoding a polypeptide having heparanase activity and expression of same in transduced cells
JOURNAL Patent: US 5968822-A 9 19-OCT-1999;

```
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match             100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
```

```
Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 2
AR080680
LOCUS       AR080680               1721 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 11 from patent US 5968822.
ACCESSION   AR080680
VERSION     AR080680.1  GI:10007410
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I. and Feinstein,E.
  TITLE     Polynucleotide encoding a polypeptide having heparanase activity
            and expression of same in transduced cells
  JOURNAL   Patent: US 5968822-A 11 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
```

ORIGIN

Query Match 100.0%; Score 1721; DB 6; Length 1721;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1721; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
```

```
Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 3
AR125603
LOCUS       AR125603                1721 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 1 from patent US 6177545.
ACCESSION   AR125603
VERSION     AR125603.1  GI:14111665
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6177545-A 1 23-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN
```

Query Match 100.0%; Score 1721; DB 6; Length 1721;
Best Local Similarity 100.0%; Pred. No. 0;

```
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721


RESULT 4
AR125604
LOCUS       AR125604               1721 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 3 from patent US 6177545.
ACCESSION   AR125604
VERSION     AR125604.1  GI:14111666
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6177545-A 3 23-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match           100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

RESULT 5
AR194189

LOCUS	AR194189	1721 bp DNA linear	PAT 20-APR-2002
DEFINITION	Sequence 1 from patent US 6348344.		
ACCESSION	AR194189		
VERSION	AR194189.1 GI:20240781		
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
	Unclassified.		
REFERENCE	1 (bases 1 to 1721)		
AUTHORS	Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M., Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Schlomi,Y.		
TITLE	Genetically modified cells and methods for expressing recombinant heparanase and methods of purifying same		
JOURNAL	Patent: US 6348344-A 1 19-FEB-2002;		
FEATURES	Location/Qualifiers		
source	1. .1721		
	/organism="unknown"		
BASE COUNT	451 a	413 c 410 g	447 t
ORIGIN			

```
  Query Match            100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
```

```
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 6
AR194190
LOCUS       AR194190                1721 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 3 from patent US 6348344.
ACCESSION   AR194190
VERSION     AR194190.1  GI:20240782
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M.,
            Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Schlomi,Y.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6348344-A 3 19-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match           100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
```

```
Qy       121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy       181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy       241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy       301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy       361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
            |||||||||||||||||||||||||||||||||||||||||
Db      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

RESULT 7
AR221285
LOCUS AR221285 1721 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 1 from patent US 6426209.
ACCESSION AR221285
VERSION AR221285.1 GI:23328256
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1721)
AUTHORS Ayal-Hershkovitz,M., Pecker,I. and Yacoby-Zeevi,O.
TITLE Genetically modified cells and methods for expressing recombinant heparanase and methods of purifying same
JOURNAL Patent: US 6426209-A 1 30-JUL-2002;
FEATURES Location/Qualifiers
source 1. .1721
/organism="unknown"
BASE COUNT 451 a 413 c 410 g 447 t
ORIGIN

Query Match 100.0%; Score 1721; DB 6; Length 1721;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1721; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy        61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy       121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
```

```
Qy       181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy       241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy       301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy       361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
            |||||||||||||||||||||||||||||||||||||||||
Db      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 8
AR221286
LOCUS       AR221286                1721 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 3 from patent US 6426209.
ACCESSION   AR221286
VERSION     AR221286.1  GI:23328257
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Ayal-Hershkovitz,M., Pecker,I. and Yacoby-Zeevi,O.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6426209-A 3 30-JUL-2002;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match           100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
```

```
Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 9
AR243203
LOCUS       AR243203               1721 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 1 from patent US 6475763.
ACCESSION   AR243203
VERSION     AR243203.1  GI:27290318
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M.,
            Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Shlomi,Y.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6475763-A 1 05-NOV-2002;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match           100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
```

```
Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 10
AR243204
LOCUS       AR243204                1721 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 3 from patent US 6475763.
ACCESSION   AR243204
VERSION     AR243204.1  GI:27290319
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Ayal-Hershkovitz,M., Moskowitz,H., Miron,D., Gilboa,A., Mimon,M.,
            Ben-Artzi,H., Yacoby-Zeevi,O., Pecker,I., Peleg,Y. and Shlomi,Y.
  TITLE     Genetically modified cells and methods for expressing recombinant
            heparanase and methods of purifying same
  JOURNAL   Patent: US 6475763-A 3 05-NOV-2002;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match           100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
```

```
Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

RESULT 11
AR287435
LOCUS AR287435 1721 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 1 from patent US 6531129.
ACCESSION AR287435
VERSION AR287435.1 GI:29725129
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1721)
AUTHORS Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
TITLE Heparanase specific molecular probes and their use in research and medical applications
JOURNAL Patent: US 6531129-A 1 11-MAR-2003;
FEATURES Location/Qualifiers
source 1. .1721
/organism="unknown"
BASE COUNT 451 a 413 c 410 g 447 t
ORIGIN

Query Match 100.0%; Score 1721; DB 6; Length 1721;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1721; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
```

```
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

RESULT 12
AR287436

```
LOCUS       AR287436                1721 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 3 from patent US 6531129.
ACCESSION   AR287436
VERSION     AR287436.1  GI:29725130
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I., Friedman,Y. and Perets,T.
  TITLE     Heparanase specific molecular probes and their use in research and
            medical applications
  JOURNAL   Patent: US 6531129-A 3 11-MAR-2003;
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unknown"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN
```

Query Match 100.0%; Score 1721; DB 6; Length 1721;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1721; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

RESULT 13
BD074427

```
LOCUS       BD074427               1721 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell.
ACCESSION   BD074427
VERSION     BD074427.1  GI:22620030
KEYWORDS    JP 2001514855-A/8.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I. and Elena,F.
  TITLE     Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell
  JOURNAL   Patent: JP 2001514855-A 8 18-SEP-2001;
            INSIGHT STRATEGY & MARKETING LTD, HADASIT MEDICAL RESEARCH SERVICES
            & DEVELOPMENT LTD
COMMENT     OS   Nucleic acid
            PN   JP 2001514855-A/8
            PD   18-SEP-2001
            PF   31-AUG-1998 JP 2000508806
            PR   02-SEP-1997 US    08/922170,02-JUL-1998 US    09/109386 PI
            IRIS PECKER,ISRAEL VLODAVSKY,FEINSTEIN ELENA
            PC   C12N15/09,A61K38/00,A61P9/10,A61P17/00,A61P29/00,A61P35/00, PC
              A61P37/00,
            PC   A61P43/00,C12N5/10,C12N9/24,C12Q1/68,G01N33/15,G01N33/50// PC
             A61K39/395,
            PC   A61K39/395,C12N15/00,A61K37/02,C12N5/00
            CC   Polynucleotide encoding polypeptide having
            heparanase activity
            CC            and
            CC   expression of the polypeptide in induced cell FH   Key
                 Location/Qualifiers
            FT   source          1. .1721
            FT                   /organism='Nucleic acid'.
FEATURES             Location/Qualifiers
     source          1. .1721
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN
```

```
  Query Match           100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
```

```
Qy       181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy       241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy       301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy       361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
            |||||||||||||||||||||||||||||||||||||||||
Db      1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
```

```
RESULT 14
BD074428
LOCUS       BD074428               1721 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell.
ACCESSION   BD074428
VERSION     BD074428.1  GI:22620031
KEYWORDS    JP 2001514855-A/9.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1721)
  AUTHORS   Pecker,I., Vlodavsky,I. and Elena,F.
  TITLE     Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell
  JOURNAL   Patent: JP 2001514855-A 9 18-SEP-2001;
            INSIGHT STRATEGY & MARKETING LTD, HADASIT MEDICAL RESEARCH SERVICES
            & DEVELOPMENT LTD
COMMENT     OS   Nucleic acid
            PN   JP 2001514855-A/9
            PD   18-SEP-2001
            PF   31-AUG-1998 JP 2000508806
            PR   02-SEP-1997 US    08/922170,02-JUL-1998 US    09/109386 PI
            IRIS PECKER,ISRAEL VLODAVSKY,FEINSTEIN ELENA
            PC   C12N15/09,A61K38/00,A61P9/10,A61P17/00,A61P29/00,A61P35/00, PC
              A61P37/00,
            PC   A61P43/00,C12N5/10,C12N9/24,C12Q1/68,G01N33/15,G01N33/50// PC
             A61K39/395,
            PC   A61K39/395,C12N15/00,A61K37/02,C12N5/00
            CC   Polynucleotide encoding polypeptide having
            heparanase activity
            CC            and
            CC   expression of the polypeptide in induced cell FH   Key
                Location/Qualifiers
            FT   CDS             (63). .(1691).
FEATURES                Location/Qualifiers
     source             1. .1721
                        /organism="unidentified"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32644"
```

```
BASE COUNT      451 a    413 c    410 g    447 t
ORIGIN

  Query Match           100.0%;  Score 1721;  DB 6;  Length 1721;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1721;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGAGCAGCCAGGTGAGCCCA 60

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721


RESULT 15
BD074430
LOCUS       BD074430               1899 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell.
ACCESSION   BD074430
VERSION     BD074430.1  GI:22620033
KEYWORDS    JP 2001514855-A/11.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1899)
  AUTHORS   Pecker,I., Vlodavsky,I. and Elena,F.
  TITLE     Polynucleotide encoding polypeptide having heparanase activity and
            expression of the polypeptide in induced cell
  JOURNAL   Patent: JP 2001514855-A 11 18-SEP-2001;
            INSIGHT STRATEGY & MARKETING LTD, HADASIT MEDICAL RESEARCH SERVICES
            & DEVELOPMENT LTD
COMMENT     OS   Nucleic acid
            PN   JP 2001514855-A/11
            PD   18-SEP-2001
            PF   31-AUG-1998 JP 2000508806
```

```
           PR   02-SEP-1997 US    08/922170,02-JUL-1998 US    09/109386 PI
           IRIS PECKER,ISRAEL VLODAVSKY,FEINSTEIN ELENA
           PC   C12N15/09,A61K38/00,A61P9/10,A61P17/00,A61P29/00,A61P35/00, PC
             A61P37/00,
           PC   A61P43/00,C12N5/10,C12N9/24,C12Q1/68,G01N33/15,G01N33/50// PC
            A61K39/395,
           PC   A61K39/395,C12N15/00,A61K37/02,C12N5/00
           CC   Polynucleotide encoding polypeptide having
           heparanase activity
           CC            and
           CC   expression of the polypeptide in induced cell FH   Key
               Location/Qualifiers
           FT   source          1. .1899
           FT                   /organism='Nucleic acid'.
FEATURES             Location/Qualifiers
     source          1. .1899
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      495 a    433 c    510 g    461 t
ORIGIN

  Query Match             99.9%;  Score 1719.4;  DB 6;  Length 1899;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1720;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTAGAGCTTTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 60
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 CTAGAGCTCTCGACTCTCCGCTGCGCGGCAGCTGGCGGGGGGGAGCAGCCAGGTGAGCCCA 238

Qy         61 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 AGATGCTGCTGCGCTCGAAGCCTGCGCTGCCGCCGCCGCTGATGCTGCTGCTCCTGGGGC 298

Qy        121 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CGCTGGGTCCCCTCTCCCCTGGCGCCCTGCCCCGACCTGCGCAAGCACAGGACGTCGTGG 358

Qy        181 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 ACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAGCCCCTCGTTCCTGTCCGTCA 418

Qy        241 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 CCATTGACGCCAACCTGGCCACGGACCCGCGGTTCCTCATCCTCCTGGGTTCTCCAAAGC 478

Qy        301 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 TTCGTACCTTGGCCAGAGGCTTGTCTCCTGCGTACCTGAGGTTTGGTGGCACCAAGACAG 538

Qy        361 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 ACTTCCTAATTTTCGATCCCAAGAAGGAATCAACCTTTGAAGAGAGAAGTTACTGGCAAT 598

Qy        421 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CTCAAGTCAACCAGGATATTTGCAAATATGGATCCATCCCTCCTGATGTGGAGGAGAAGT 658

Qy        481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAGAAAAAGT 718

Qy        541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTTGCAAACTGCT 778

Qy        601 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 CAGGACTGGACTTGATCTTTGGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAGTGGA 838

Qy        661 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 ACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCTTCCAAGGGGTATAACATTTCTT 898

Qy        721 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GGGAACTAGGCAATGAACCTAACAGTTTCCTTAAGAAGGCTGATATTTTCATCAATGGGT 958

Qy        781 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CGCAGTTAGGAGAAGATTATATTCAATTGCATAAACTTCTAAGAAAGTCCACCTTCAAAA 1018

Qy        841 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATGCAAAACTCTATGGTCCTGATGTTGGTCAGCCTCGAAGAAAGACGGCTAAGATGCTGA 1078

Qy        901 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AGAGCTTCCTGAAGGCTGGTGGAGAAGTGATTGATTCAGTTACATGGCATCACTACTATT 1138

Qy        961 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 TGAATGGACGGACTGCTACCAGGGAAGATTTTCTAAACCCTGATGTATTGGACATTTTTA 1198

Qy       1021 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 TTTCATCTGTGCAAAAAGTTTTCCAGGTGGTTGAGAGCACCAGGCCTGGCAAGAAGGTCT 1258

Qy       1081 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 GGTTAGGAGAAACAAGCTCTGCATATGGAGGCGGAGCGCCCTTGCTATCCGACACCTTTG 1318

Qy       1141 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 CAGCTGGCTTTATGTGGCTGGATAAATTGGGCCTGTCAGCCCGAATGGGAATAGAAGTGG 1378

Qy       1201 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 TGATGAGGCAAGTATTCTTTGGAGCAGGAAACTACCATTTAGTGGATGAAAACTTCGATC 1438

Qy       1261 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 CTTTACCTGATTATTGGCTATCTCTTCTGTTCAAGAAATTGGTGGGCACCAAGGTGTTAA 1498

Qy       1321 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 TGGCAAGCGTGCAAGGTTCAAAGAGAAGGAAGCTTCGAGTATACCTTCATTGCACAAACA 1558

Qy       1381 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 CTGACAATCCAAGGTATAAAGAAGGAGATTTAACTCTGTATGCCATAAACCTCCATAACG 1618

Qy       1441 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 TCACCAAGTACTTGCGGTTACCCTATCCTTTTTCTAACAAGCAAGTGGATAAATACCTTC 1678

Qy       1501 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TAAGACCTTTGGGACCTCATGGATTACTTTCCAAATCTGTCCAACTCAATGGTCTAACTC 1738

Qy       1561 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 TAAAGATGGTGGATGATCAAACCTTGCCACCTTTAATGGAAAAACCTCTCCGGCCAGGAA 1798

Qy       1621 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 GTTCACTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGTGATAAGAAATGCCAAAGTTG 1858

Qy       1681 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1721
              |||||||||||||||||||||||||||||||||||||||||
Db       1859 CTGCTTGCATCTGAAAATAAAATATACTAGTCCTGACACTG 1899

Search completed: February 16, 2004, 11:42:58
Job time : 8222.88 secs
```